

Stable fixed points of combinatorial threshold-linear networks

Carina Curto, Jesse Geneson, Katherine Morrison

November 18, 2023

Abstract. Combinatorial threshold-linear networks (CTLNs) are a special class of recurrent neural networks whose dynamics are tightly controlled by an underlying directed graph. Recurrent networks have long been used as models for associative memory and pattern completion, with stable fixed points playing the role of stored memory patterns in the network. In prior work, we showed that *target-free cliques* of the graph correspond to stable fixed points of the dynamics, and we conjectured that these are the only stable fixed points possible [1, 2]. In this paper, we prove that the conjecture holds in a variety of special cases, including for networks with very strong inhibition and graphs of size $n \leq 4$. We also provide further evidence for the conjecture by showing that sparse graphs and graphs that are nearly cliques can never support stable fixed points. Finally, we translate some results from extremal combinatorics to obtain an upper bound on the number of stable fixed points of CTLNs in cases where the conjecture holds.

Keywords: stable fixed points, attractor neural networks, threshold-linear networks, cliques, Collatz-Wielandt formula

Contents

1	Introduction	2
1.1	Threshold-linear networks	2
1.2	Fixed points of TLNs	3
1.3	The CTLN model and stable fixed points of CTLNs	4
1.4	Early evidence: conjecture holds for symmetric graphs and oriented graphs . .	6
1.5	Summary of main results	7
1.6	Bounds on the number of stable fixed points	8
1.7	Discussion	9
2	Preliminaries	11
2.1	General background on fixed points and stability	11
2.2	Uniform in-degree and simply-added splits	12
2.3	Techniques for ruling out stable motifs	13
3	Degree bounds and proofs of Theorems 1.7 and 1.9	16
4	Near-cliques and proof of Theorem 1.8	18
5	Composite graphs that are never stable motifs	20
6	Stable motifs up to size 4 and proof of Theorem 1.6	22
7	Enumerating target-free cliques and proof of Theorem 1.10	26

1. Introduction

One of the major challenges in theoretical neuroscience is to relate network connectivity to function. In the study of attractor neural networks, dating back at least to the Hopfield model [3, 4], the attractors of interest have typically been stable fixed points of the dynamics. These fixed points represent stored memory patterns, and the function of the network is to transform an input pattern into a stable fixed point via the evolution of network dynamics. In this way, attractor neural networks can perform pattern completion, image classification, and memory retrieval [5, 6, 7]. Stable fixed points have also been used to model sensory representations, such as orientation selectivity and surround suppression in visual cortex [8, 9], as well as position coding in hippocampus [10], perceptual bistability [11], and decision-making [12]. How does a network encode a set of attractors? More specifically, given a network connectivity matrix W , what can we say about the stable fixed points?

In this work, we address this question in the context of combinatorial threshold-linear networks (CTLNs). These are a special case of threshold-linear networks (TLNs), which are recurrent network models that have been commonly used in computational neuroscience for decades [13, 14, 15, 5, 16, 17, 18, 6, 19]. In the case of CTLNs, the connectivity matrix, W , is tightly controlled by a directed graph, G , leading to a fixed point structure that is closely related to properties of G . In prior work, we showed that certain cliques of G , called *target-free cliques*, correspond to stable fixed points of the dynamics. Furthermore, we conjectured that these are the only stable fixed points possible [1, 2]. Here we prove that the conjecture holds in a variety of special cases, including for networks with very strong inhibition and graphs of size $n \leq 4$. We also provide further evidence for the conjecture by showing that sparse graphs and graphs that are nearly cliques can never support stable fixed points. Finally, we translate some results from extremal combinatorics to obtain an upper bound on the number of stable fixed points of CTLNs in cases where the conjecture holds.

We begin with some basic background on TLNs and CTLNs, enough so that we can state the main results.

1.1. Threshold-linear networks

Briefly, a TLN is a rate model consisting of n nodes, with dynamics governed by the system of ordinary differential equations:

$$\frac{dx_i}{dt} = -x_i + \left[\sum_{j=1}^n W_{ij}x_j + b_i \right]_+, \quad i = 1, \dots, n. \quad (1)$$

The dynamic variables $x_1, \dots, x_n$ give the activity levels¹ of nodes $1, \dots, n$. The matrix entries W_{ij} are directed connection strengths between pairs of nodes, the vector $b = (b_1, \dots, b_n) \in \mathbb{R}^n$ represents the external drive to each node, and the threshold-nonlinearity $[\cdot]_+$ is given by $[y]_+ = \max\{y, 0\}$.

TLNs are a particularly appealing model because they exhibit the full range of nonlinear behaviors, such as multistability, periodic attractors, and chaos, while remaining relatively

¹ If the nodes are neurons, the activity level is typically called a ‘firing rate.’

mathematically tractable because of their piecewise linear dynamics. *Multistability*, the co-existence of multiple stable fixed points in a network, has made TLNs particularly useful as models of associative memory storage and retrieval, similar to the Hopfield model [3]. In this context, the stable fixed points represent the set of memories encoded in the network. The network can then perform a sort of *pattern completion*: the activity evolves from a given initial condition to converge to one of the multiple stored patterns of the network. The first major developments in the mathematical theory of TLNs were made for *symmetric networks*, with a focus on characterizing the stable fixed points [15, 5, 16].

1.2. Fixed points of TLNs

Recall that a fixed point is a point $x^* \in \mathbb{R}^n$ that satisfies $dx_i/dt|_{x=x^*} = 0$ for each $i \in \{1, \dots, n\}$. The *support* of a fixed point is the subset of active nodes,

$$\text{supp } x^* \stackrel{\text{def}}{=} \{i \mid x_i^* > 0\}.$$

We restrict to considering TLNs that are *nondegenerate* (see Section 2.1 for the precise definition). One consequence of the nondegeneracy condition (that is generically satisfied) is that there is at most one fixed point per support [2]. We can thus label all the fixed points of a given network by their supports. We denote this as:

$$\text{FP}(W, b) \stackrel{\text{def}}{=} \{\sigma \subseteq [n] \mid \sigma = \text{supp } x^* \text{ for some fixed pt } x^* \text{ of the associated TLN}\},$$

where $[n] \stackrel{\text{def}}{=} \{1, \dots, n\}$. Note that for each support $\sigma \in \text{FP}(W, b)$, the fixed point itself is easily recovered. Outside the support, $x_i^* = 0$ for all $i \notin \sigma$. Within the support, x^* is given by:

$$x_\sigma^* = (I - W_\sigma)^{-1} b_\sigma, \tag{2}$$

where x_σ^* and b_σ are column vectors obtained by restricting x^* and b to the indices in σ , and W_σ is the induced principal submatrix obtained by restricting rows and columns of W to σ .

Recall that a fixed point x^* is *stable* if small perturbations of a solution from that point decay over time; consequently, within a neighborhood of the point, all trajectories converge to that fixed point. Otherwise, the fixed point is said to be *unstable*. The following result shows that stability can be checked directly from the matrix W ; this was first proven for symmetric W in [16], and later generalized to arbitrary W in [17].

Theorem 1.1 ([16, 17]). *Let (W, b) be a TLN on n nodes, and consider $\sigma \in \text{FP}(W, b)$. Then the fixed point of (1) with support σ is stable if and only if the principal submatrix $(-I + W)_\sigma$ is a stable matrix – i.e., if all its eigenvalues have negative real part.*

Significant work has been done characterizing the collection of stable fixed point supports of symmetric W through the lens of *permitted sets*, initially in [16] and extended in [17, 18, 6]. These advances were largely confined to the case of symmetric W because there is more machinery for analyzing eigenvalues in that setting, such as Cauchy’s Interlacing Theorem and results on nondegenerate square distance matrices. Thus, to make progress in the nonsymmetric case, it is useful to restrict to a different class of networks that extend well beyond the symmetric case, but maintain mathematical tractability.

1.3. The CTLN model and stable fixed points of CTLNs

Combinatorial threshold-linear networks (CTLNs) are a special class of threshold-linear networks, first introduced in [1], whose dynamics are tightly controlled by an underlying directed graph. CTLNs have uniform inputs, $b_i = \theta$, and a connectivity matrix W that is fully determined by a simple² directed graph G and continuous parameters ε and δ . Specifically $W = W(G, \varepsilon, \delta)$ has the form:

$$W_{ij} = \begin{cases} 0 & \text{if } i = j, \\ -1 + \varepsilon & \text{if } j \rightarrow i \text{ in } G, \\ -1 - \delta & \text{if } j \not\rightarrow i \text{ in } G. \end{cases} \quad (3)$$

where $j \rightarrow i$ indicates the presence of an edge from j to i in the graph G , while $j \not\rightarrow i$ indicates the absence of such an edge. We additionally require that $\theta > 0$, $\delta > 0$, and $0 < \varepsilon < \frac{\delta}{\delta+1}$; or equivalently, $0 < \varepsilon < 1$ and $\delta > \frac{\varepsilon}{1-\varepsilon}$. When these conditions are met, we say that the parameters are within the *legal range*³ (see Figure 1B).

The simplicity of binary synapses has made the CTLN model particularly mathematically tractable, while enabling us to explore networks beyond the symmetric case. Moreover, the tight connection between the connectivity graph G and the matrix W has allowed us to isolate the role of connectivity in shaping dynamics. This has led to the development of a series of *graph rules* that directly connect fixed points of the network to particular features or subgraphs of G [1, 2, 20, 21]. The primary focus of this paper is a graphical characterization of the stable fixed points of CTLNs. This requires some terminology: A *clique* is subgraph in which there are all-to-all bidirectional connections. We say that a node $k \notin \sigma$ is a *target* of σ if $i \rightarrow k$ for every $i \in \sigma$. We refer to cliques that do not have any targets in G as *target-free cliques*.

In [1], it was shown that every target-free clique of G is the support of a stable fixed point of any CTLN with graph G , irrespective of parameters (within the legal range). Moreover, these appeared to be the only graph structures that gave rise to stable fixed points. These observations led to the following conjecture.

Conjecture 1.2 ([1]). *Let $W = W(G, \varepsilon, \delta)$ be a CTLN on n nodes with graph G , and let $\sigma \subseteq [n]$. Then σ is the support of a stable fixed point if and only if σ is a target-free clique.*

As an illustration of the conjecture, consider the CTLN in Figure 2. The collection of fixed point supports, $\text{FP}(W)$, are given below the graph of the network, with the stable fixed points in bold.⁴ Note that the supports of the stable fixed points are precisely the target-free cliques 48 and 189. The other cliques in the graph all have targets and consequently do not support fixed points of the network (see Theorem 2.5); for example, the clique 78 has node 1 as a target, and $78 \notin \text{FP}(W)$. Also, notice that no other type of subgraph supports a stable fixed point other than the target-free cliques.

Figure 2B shows the attractors of the CTLN, which include two dynamic attractors in addition to the stable fixed point attractors noted above. Note that the high-firing nodes

²A directed graph is *simple* if it does not have self-loops or multiple edges between a pair of nodes

³Within this parameter range, $W_{ij} \leq 0$ which guarantees that the network is competitive and activity is bounded. The particular constraint on ε was motivated by a result in [1] to guarantee that a unidirectional edge is always unstable.

⁴For compactness, we write ijk to denote the subset $\{i, j, k\}$.

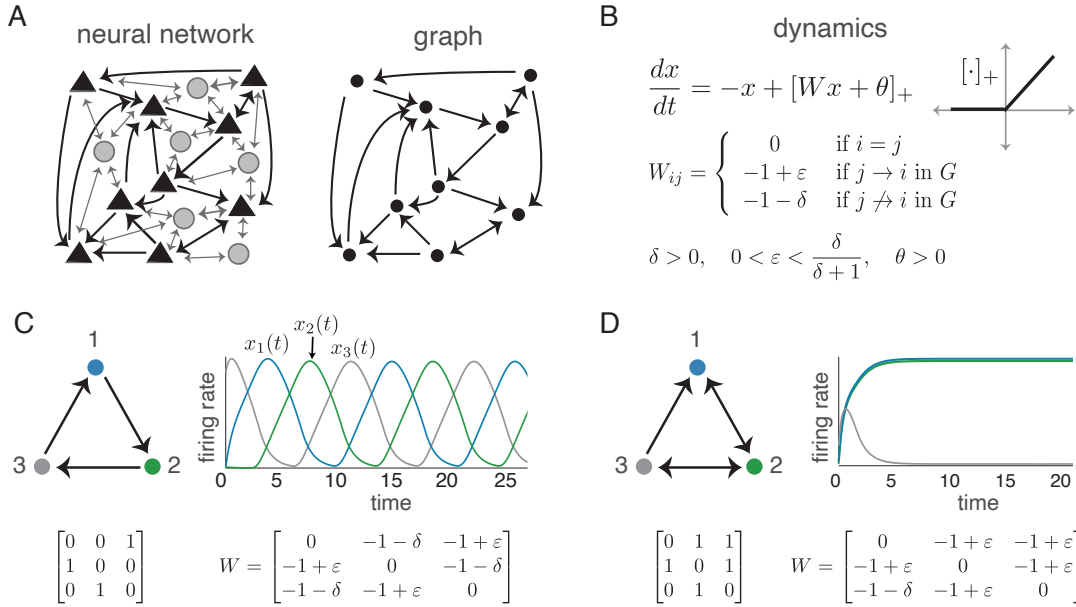


Figure 1: **Combinatorial threshold-linear networks with strong and weak inhibition** (modeling excitatory neurons in a sea of inhibition). (A) (Left) A neural network with excitatory pyramidal neurons (triangles) and a background network of inhibitory interneurons (gray circles) that produce a global inhibition. (Right) The graph of the network retains only the excitatory neurons and the connections between them. In the corresponding CTLN, the arrows in the graph indicate weak inhibition from the sum of excitation and a global background inhibition. The absence of an edge thus indicates strong inhibition. (B) The equations for a CTLN network. (C) (Left) The 3-cycle graph with its corresponding adjacency matrix and W matrix below. (Right) Network activity follows the arrows in the graph, with peak activity occurring sequentially in the cyclic order 123. (D) A CTLN with one stable fixed point, which has support $\{1, 2\}$ (network activity shown on right). Note that $\{1, 2\}$ is a *target-free* clique. The clique $\{2, 3\}$ does not have a corresponding fixed point; node 1 is a target of this clique. All simulations have parameters $\varepsilon = 0.25$, $\delta = 0.5$, and $\theta = 1$.

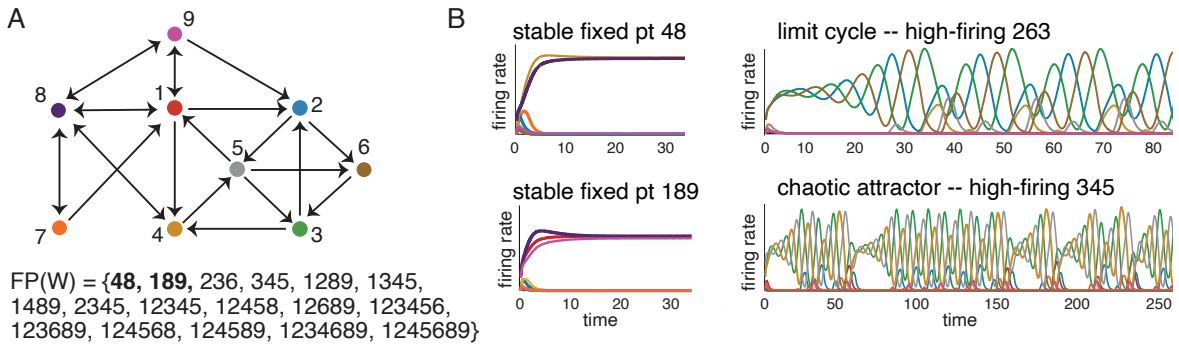


Figure 2: **A CTLN with multiple fixed points and attractors.** (A) Graph of a CTLN together with its fixed point supports $\text{FP}(W) \stackrel{\text{def}}{=} \text{FP}(W(G, \varepsilon, \delta), \theta)$ for $\varepsilon = 0.25$, $\delta = 0.5$, $\theta = 1$. The supports of stable fixed points are bolded. (B) The network has four attractors: two stable fixed points, one limit cycle, and one chaotic attractor. The equations for the dynamics are identical in each case; only the initial conditions differ, and these determine which attractor the solution converges to.

in each of these dynamic attractors correspond to the unstable fixed points with minimal supports 236 and 345. These minimal fixed points come from special types of subnetworks, known as *core motifs*, which have been observed computationally to give rise to attractors of CTLNs [22]. Target-free cliques are provably core motifs, and we conjecture these are the

only core motifs that correspond to stable fixed points.

The remainder of this paper is focused on collecting evidence towards proving Conjecture 1.2. To state these results, we first review some useful graph-theoretic terminology.

Graph theory terminology. Let G be a simple directed graph on n nodes and $\sigma \subseteq [n]$. We denote by $G|_\sigma$ the induced subgraph obtained by restricting to the nodes in σ and keeping only the edges between nodes in σ . Recall that a subgraph $G|_\sigma$ is a *clique* if every pair of nodes in σ has a bidirectional edge between them. As an abuse of notation, when G is given, we may refer to a subset σ as being a clique to indicate that $G|_\sigma$ is a clique. A clique is called *maximal* if it is not properly contained in any larger clique. Note that every *target-free clique* must be maximal, since any larger clique containing it would have a node $k \notin \sigma$ that is a target of σ . As an illustration of these terms, consider the graph in Figure 2A. We see that 189 is a target-free clique, and thus is necessarily maximal. The subsets 18, 19, and 89 are all non-maximal cliques, and thus have targets. Finally, 78 is a maximal clique that is not target-free: node 1 is a target.

As a generalization of cliques, we also consider directed cliques. We say that $G|_\sigma$, or equivalently σ , is a *directed clique* if there exists an ordering of the nodes $1, \dots, |\sigma|$ such that $i \rightarrow j$ in $G|_\sigma$ whenever $i < j$. Note that there are no constraints on the back edges $j \rightarrow i$ when $i < j$, and thus directed cliques include cliques as a special case. In Figure 2A, the subgraphs $G|_{256}$, $G|_{129}$, and $G|_{148}$ are all examples of directed cliques.

A graph is *oriented* if it has no bidirectional edges, and thus no cliques of size greater than one. Figure 1C shows an example oriented graph, the 3-cycle. As another example, for the graph G in Figure 2A, $G|_{\{1, \dots, 6\}}$ is oriented. A *sink* is a node with no outgoing edges in G . Note that singletons are trivially cliques, and sinks are precisely the target-free cliques of size 1.

Finally, we will also be interested in uniform in-degree graphs. We say that $G|_\sigma$, or equivalently σ , is *uniform in-degree* if all nodes in the graph $G|_\sigma$ have the same number of incoming edges from nodes in σ . Note that the in-degrees need not agree in the full graph G , only within the restricted subgraph $G|_\sigma$. There are no constraints on the out-degrees.

1.4. Early evidence: conjecture holds for symmetric graphs and oriented graphs

Conjecture 1.2 was motivated in part by the fact that it had previously been proven to hold in two extreme cases: when the graph G is *symmetric*⁵, so that it only contains bidirectional edges, and when G is oriented, so that it has no bidirectional edges.

Theorem 1.3 ([6]). *Let G be a symmetric graph. Then for any CTLN with graph G , the supports of the stable fixed points are precisely the maximal cliques of G .*

This theorem establishes that Conjecture 1.2 holds for symmetric G because in a symmetric graph, the target-free cliques are precisely the maximal cliques. To see this, recall that

⁵We say that a directed graph is *symmetric* if between any pair of nodes there is either a bidirectional edge or no edge.

in any graph a target-free clique is necessarily maximal. On the other hand, maximal cliques in symmetric graphs must be target-free because any target $k \notin \sigma$ receiving from all nodes in σ would have bidirectional edges to all nodes in σ , contradicting the fact that σ is a maximal clique.

At the other extreme, consider an oriented graph G . Since G has no bidirectional edges, it contains no cliques of size greater than 1. Thus the only cliques in G are the singletons, and a singleton is target-free precisely when it has no outgoing edges, i.e. it is a *sink*. Thus, in the case of oriented graphs, the conjecture reduces to the following theorem from [1].

Theorem 1.4 ([1]). *Let G be an oriented graph. Then for any CTLN with graph G , the supports of the stable fixed points are precisely the singletons that are sinks in G .*

Remark 1.5. A key ingredient to the proof of Theorem 1.4 in [1] was the requirement that $\varepsilon < \frac{\delta}{\delta+1}$, which defines the legal parameter range of CTLNs. Without this condition, there may be nontrivial subgraphs of oriented graphs that can support stable fixed points. For example, it is straightforward to check that for the 3-cycle (the oriented graph in Figure 1C), the fixed point with support 123 is stable whenever $\varepsilon > \delta$.

These extreme cases provide evidence for Conjecture 1.2, but we are still left with the question of whether the conjecture holds more generally. Are there other types of graphs where we can guarantee that target-free cliques are the *only* subgraphs that support stable fixed points? Can we prove that certain types of subgraphs *never* support stable fixed points, irrespective of what graph they are embedded in? These questions are the primary focus of this paper. In the following section, we summarize the main results we have proven in this vein, which provide further evidence in support of Conjecture 1.2.

1.5. Summary of main results

To state the results, we first set some notation. We assume throughout the remainder of the paper that we are always considering a nondegenerate CTLN $W = W(G, \varepsilon, \delta)$ with arbitrary parameters within the legal range (unless otherwise noted). Thus the results are parameter-independent unless explicitly noted otherwise. The graph G is assumed to have n nodes, and $\sigma \subseteq [n]$. Finally, as shorthand, we often refer to σ as having a particular graphical property, when technically we mean that the subgraph $G|_{\sigma}$ has that property.

Our first major result shows that among small subgraphs, the only ones that can support stable fixed points are target-free cliques. Consequently if there are counterexamples to the conjecture, they can only occur in subgraphs of size at least 5.

Theorem 1.6. *If $|\sigma| \leq 4$, then σ supports a stable fixed point of W if and only if σ is a target-free clique in G .*

The proof of Theorem 1.6 is postponed to Section 6, as it relies on the application of many of the other main results to rule out families of subgraphs as possible supports of stable fixed points. One such collection of graphs that we can rule out are those that are very far from having clique-like structure by way of being relatively sparse.

Theorem 1.7. *If $G|_\sigma$ has maximum in-degree $d_{\max}^{\text{in}} < \frac{|\sigma|}{2}$, then σ cannot support a stable fixed point.*

At the other extreme, Theorem 1.8 shows that various dense graphs that are “near-cliques” also cannot support stable fixed points. Recall that *directed cliques* are graphs that generalize clique structure in directed graphs (see the graph theory terminology in Section 1.3 for the precise definition).

Theorem 1.8. *Let σ be a directed clique or contain a clique of size $|\sigma| - 1$. Then σ cannot support a stable fixed point unless σ is a clique.*

Theorem 1.7 is proven in Section 3, while the proof of Theorem 1.8 is given in Section 4. In Section 5, we examine infinite families of graphs, known as *composite graphs*, that are built from connecting arbitrary component subgraphs in a prescribed manner. We show that when the component subgraphs have certain properties, e.g., sparsity, then the full graph is guaranteed to never support a stable fixed point. On the flip side, when the component subgraphs are dense, but there are relatively few edges between components, then again the full graph is guaranteed never to support a stable fixed point.

The results presented so far show that the conjecture holds for certain families of graphs, irrespective of the CTLN parameters. In contrast, the following theorem shows that if we restrict to sufficiently strong inhibition, i.e. sufficiently large δ , then the conjecture holds for all graphs, provided the CTLN lies within this parameter regime.

Theorem 1.9. *Let G be an arbitrary graph and $W = W(G, \varepsilon, \delta)$ be a corresponding CTLN whose parameters satisfy*

$$\delta > \frac{\varepsilon}{1 - \varepsilon}(n^2 - n - 1).$$

Then for any $\sigma \subseteq [n]$, σ supports a stable fixed point of W if and only if σ is a target-free clique in G .

The condition on δ in Theorem 1.9 is rather extreme, since it grows with n^2 ; Theorem 3.3 in Section 3 provides more generous bounds on δ when the size of σ or the maximum in-degree of $G|_\sigma$ can be taken into account. In fact, Theorem 1.7, which holds for the full legal parameter range, is a corollary of this result. In each of these cases, though, the requirement of large δ is an artifact of the techniques used to prove the results; we do not believe that this restriction is actually necessary for the conjecture to hold. Moreover, we present these results as proof of concept that there exists a parameter regime where the conjecture holds for all graphs, not only for restricted families of graphs with special properties.

1.6. Bounds on the number of stable fixed points

In previous work, it was shown that the maximum number of stable fixed points of a CTLN is at most 2^{n-1} (Corollary 1 in [2]). This bound was derived from constraints on the vector field that follow from the Poincaré-Hopf theorem; it does not exploit any graphical structure that might constrain how the subgraphs supporting stable fixed points can coexist within a

graph. Here we show that there is a tighter bound on the number of target-free cliques that can coexist in a graph on n nodes. Thus, if the conjecture is true, or in any context where it holds, this provides a tighter upper bound on the number of stable fixed points of a CTLN.

To derive this bound, we make use of a straightforward correspondence between target-free cliques of directed graphs and maximal cliques of undirected graphs (see Section 7 for more details). This enables us to exploit results on maximal cliques, which have been significantly studied in extremal graph theory. Applying a result from Moon and Moser [23], we immediately obtain the following upper bound on the number of target-free cliques in a directed graph.

Theorem 1.10. *The maximum number of target-free cliques in a directed graph of size n is*

$$\max \# \text{ of target-free cliques} = \begin{cases} 3^{n/3} & \text{if } n \equiv 0 \pmod{3} \\ 4 \cdot 3^{\lfloor n/3 \rfloor - 1} & \text{if } n \equiv 1 \pmod{3} \\ 2 \cdot 3^{\lfloor n/3 \rfloor} & \text{if } n \equiv 2 \pmod{3}. \end{cases}$$

Moreover, it is straightforward to construct the graph of a CTLN that attains this upper bound. Figure 3 illustrates the construction in the case when $n \equiv 0 \pmod{3}$. The graph is comprised of $n/3$ component subgraphs, each containing 3 nodes with no edges between them (an *independent set*). The components are connected following a *clique union*⁶ architecture, in which there is a bidirectional edge between every pair of nodes in different components. This graph contains $3^{n/3}$ maximal cliques of size $n/3$, each consisting of one node per component. To see that each of these cliques is target-free, let σ be such a clique and consider any $k \notin \sigma$. The node k must be in some G_i , and since σ contains a node from every component subgraph, it must contain a j in G_i . Since G_i is an independent set, $j \not\rightarrow k$, and so k cannot be a target of σ . Thus, σ is target-free. To modify this construction for the case when $n \equiv 1 \pmod{3}$, we simply make one of the component subgraphs contain 4 nodes instead of 3. For the case when $n \equiv 2 \pmod{3}$, we add an additional component subgraph that is an independent set on 2 nodes.

1.7. Discussion

A target-free clique σ is always minimal in $\text{FP}(W)$ with respect to inclusion because any proper subset of σ is necessarily a clique with a target, and thus cannot be in $\text{FP}(W)$ (see Theorem 2.5). Moreover, target-free cliques are a special case of what are known as core motifs,⁷ which have been observed computationally to have a close correspondence with the attractors of CTLNs [22]. If Conjecture 1.2 is true, then all stable fixed points are guaranteed to be core motifs, and thus minimal in $\text{FP}(W)$. Even if the conjecture is not true, it may still be the case that all stable fixed points of CTLNs are core motifs or at least minimal. This may be considered a weaker version of Conjecture 1.2.

⁶Clique unions are a special type of *composite graph*, first introduced in [2], and further considered here in Section 5.

⁷Note that σ is a *core motif* if it is the unique fixed point of its restricted subnetwork, i.e. $\text{FP}(W_\sigma) = \{\sigma\}$. For any $\tau \subsetneq \sigma$, we see that $\tau \notin \text{FP}(W_\sigma)$ and consequently $\tau \notin \text{FP}(W)$ (see [2, Corollary 2] for more details). Thus, core motifs are always minimal supports within $\text{FP}(W)$.

It is already known that Conjecture 1.2 only applies to CTLNs, as there are TLNs that violate the conjecture in both directions: there are TLNs where (1) a target-free clique of the corresponding graph does not support a stable fixed point, and others where (2) there are stable fixed points supported on other types of subgraphs. However, it is still an open question whether stable fixed points always correspond to core motifs. In particular, the following questions remain.

Question. For a general TLN (W, b) , if $\sigma \in \text{FP}(W, b)$ supports a stable fixed point, is σ guaranteed to be a core motif? If not, is σ at least guaranteed to be minimal in $\text{FP}(W, b)$?

We conjecture that the answer to this question is yes, stable fixed points are minimal fixed points. However, we do not currently have strong evidence in support of this outside of the case of CTLNs, and so it requires further exploration.

The remainder of the paper is organized as follows. Section 2 reviews general background on fixed points of CTLNs and then lays out the key techniques to prove a subset cannot support a stable fixed point, which will be used throughout the following sections to prove the main results. Section 3 is devoted to proving Theorem 1.7, guaranteeing that sparse graphs never support stable fixed points, and Theorem 1.9, showing the conjecture holds within a subset of the legal parameter range. We will see that both of these results are in fact immediate corollaries of a single technical theorem that connects $|\sigma|$ and the maximum in-degree of $G|_\sigma$ to regions of parameter space in which a fixed point supported on σ is guaranteed to be unstable. Section 4 collects various results about “near-cliques” and concludes with the proof of Theorem 1.8. In Section 5, we consider a broad class of *composite graphs*, which include clique unions as a special case. We prove that sufficiently sparse composite graphs never support a stable fixed point. Next, Section 6 considers all possible fixed point supports up to size 4 and shows that, among subgraphs that can support fixed points, cliques are the only ones with corresponding fixed points that are stable, proving Theorem 1.6. Finally, in Section 7, we develop the correspondence between target-free cliques of directed

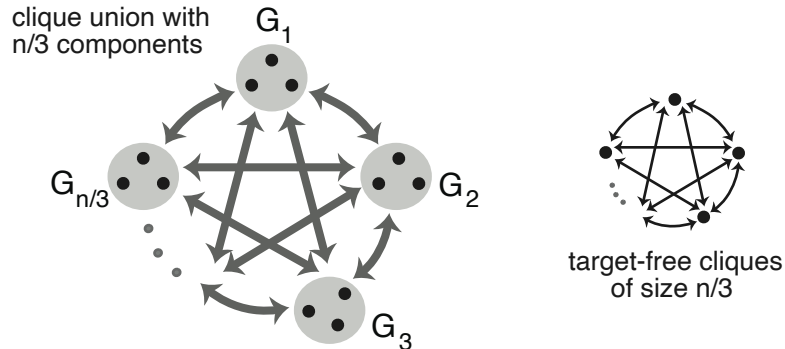


Figure 3: **Graph with maximum number of target-free cliques.** (Left) Cartoon of graph on n nodes, where $n \equiv 0 \pmod{3}$, that is a *clique union* of component subgraphs $G_1, \dots, G_{n/3}$, each of which is an independent set on 3 nodes. Thick edges indicate that every node in one component send edges to every node in the other component, so there are all-to-all connections between all nodes in different components. (Right) One of the $3^{n/3}$ maximal cliques from the graph on the left. Each such clique is also target-free.

graphs and maximal cliques of undirected graphs, and provide a summary of the literature on the latter to obtain counts of target-free cliques and algorithms for enumerating them.

2. Preliminaries

2.1. General background on fixed points and stability

Throughout this work, we will be focused on the subsets σ that can support fixed points of a CTLN $W = W(G, \varepsilon, \delta)$, i.e. subsets $\sigma \in \text{FP}(W)$ where

$$\text{FP}(W) \stackrel{\text{def}}{=} \{\sigma \subseteq [n] \mid \sigma \text{ is the support of a fixed point}\}.$$

Note that we omit θ from the notation $\text{FP}(W)$ because the value of $\theta > 0$ has no impact on the fixed point supports, it simply scales the precise value of the corresponding fixed points. To exploit previous characterizations of fixed points in terms of their supports [2], we will restrict consideration to CTLNs that are nondegenerate, as defined below.

Definition 2.1. We say that a CTLN (W, θ) is *nondegenerate* if

- $\det(I - W_\sigma) \neq 0$ for each $\sigma \subseteq [n]$, and
- for each $\sigma \subseteq [n]$ and all $i \in \sigma$, the corresponding Cramer's determinant is nonzero: $\det((I - W_\sigma)_i; \theta) \neq 0$.

Note that almost all CTLNs are nondegenerate, since having a zero determinant is a highly fine-tuned condition. The notation $\det(A_i; b)$ denotes the determinant obtained by replacing the i^{th} column of A with the vector b , as in Cramer's rule. In the case of a restricted matrix, $((A_\sigma)_i; b_\sigma)$ denotes the matrix obtained from A_σ by replacing the column corresponding to the index $i \in \sigma$ with b_σ (note that this is not typically the i^{th} column of A_σ).

In [2], multiple characterizations of $\text{FP}(W)$ were developed for nondegenerate threshold-linear networks in general as well as CTLNs specifically, including a variety of *graph rules* for CTLNs. As an immediate consequence of one of these characterizations, it was shown that σ is the support of a fixed point, i.e. $\sigma \in \text{FP}(W)$, precisely when

- (1) $\sigma \in \text{FP}(W_\sigma)$, and
- (2) $\sigma \in \text{FP}(W_{\sigma \cup \{k\}})$ for every $k \notin \sigma$

(see [2, Corollary 2] for more details). We say that σ is a *permitted motif* of W when it is a fixed point of its restricted subnetwork, so that condition (1) holds. This condition is satisfied whenever the fixed point corresponding to σ has strictly positive entries within σ , i.e. when $(I - W_\sigma)^{-1} \mathbf{1}_\sigma > 0$ (see Equation (2)).

Given a permitted motif σ , we say that σ *survives* to support a fixed point in the full network when condition (2) is satisfied. Note that whether a subset σ is permitted depends only on the subgraph $G|_\sigma$ (and potentially the choice of parameters ε and δ), while its survival will depend on the embedding of this subgraph in the full graph.

We say that σ is a *stable motif* if it is a permitted motif and its corresponding fixed point is stable, which occurs when all the eigenvalues of $-I + W_\sigma$ have negative real part (or equivalently all the eigenvalues of $I - W_\sigma$ have positive real part) by Theorem 1.1. Thus, stable motifs are the only candidate subsets for supporting stable fixed points, but whether they actually survive to yield stable fixed points of the full network will depend on their embedding.

Definition 2.2. Let (W, θ) be a CTLN on n neurons, and let $\sigma \subseteq [n]$. We say that σ is a *stable motif* of the network if the following two conditions hold:

- (i) $(I - W_\sigma)^{-1}1_\sigma > 0$, and
- (ii) $-I + W_\sigma$ is a stable matrix (i.e., all eigenvalues have negative real part).

Again, note that σ being a stable motif is a necessary, but not sufficient, condition for the existence of a stable fixed point with support σ in the full network (W, θ) ; one must also check that this stable fixed point actually *survives* in the full network.

2.2. Uniform in-degree and simply-added splits

As our first example of graphs that are guaranteed to be permitted motifs, and thus candidate stable motifs, we turn to those with *uniform in-degree*.

Definition 2.3. We say that σ has *uniform in-degree* d if every node in σ has d incoming edges within $G|_\sigma$, i.e. if the in-degree $d_i^{\text{in}} = d$ for all $i \in \sigma$.

Theorem 2.4 (Theorem 5 (uniform in-degree) in [2]). *Suppose σ has uniform in-degree d in a graph G . Then σ is a permitted motif of any CTLN W with graph G .*

For $k \notin \sigma$, let $d_k \stackrel{\text{def}}{=} |\{i \in \sigma \mid i \rightarrow k\}|$ be the number of edges k receives from σ . Then

$$\sigma \in \text{FP}(W) \Leftrightarrow d_k \leq d \text{ for all } k \notin \sigma.$$

Furthermore, if $d < |\sigma|/2$ and $|\sigma| > 1$, then the fixed point is unstable. If $d = |\sigma| - 1$, i.e. if σ is a clique, then the fixed point is stable.

The particular proof techniques used for Theorem 2.4 only enabled us to prove that σ is unstable when $d < |\sigma|/2$, but we conjecture that this holds whenever $d < |\sigma| - 1$, i.e. whenever σ is not a clique.

Combining Theorem 2.4 with a straightforward computation of eigenvalues yields the following result characterizing cliques.

Theorem 2.5 ([2]). *Let σ be a clique in G and let W be a CTLN with graph G . Then*

$$\sigma \in \text{FP}(W) \Leftrightarrow \sigma \text{ is target-free.}$$

Moreover, σ is a stable motif, with eigenvalues of $I - W_\sigma$ given by

$$\text{eig}(I - W_\sigma) = \{\varepsilon, |\sigma|(1 - \varepsilon) - \varepsilon\},$$

where ε has multiplicity $|\sigma| - 1$.

One graph structure that will prove particularly useful for getting a handle on eigenvalues of larger motifs is that of *simply-added splits*. In particular, if a motif has a simply-added split containing a uniform in-degree subgraph, then the eigenvalues of the uniform in-degree subgraph will be inherited as eigenvalues of the full motif (Lemma 2.6).

Suppose $\sigma = \tau \dot{\cup} \omega$ with $\tau, \omega \neq \emptyset$ and $\tau \cap \omega = \emptyset$. We say that ω is *simply-added* onto τ (or equivalently that σ has a *simply-added split* $\tau \dot{\cup} \omega$) if each node in ω treats all the nodes in τ identically in terms of its outgoing edges, i.e. for each $i \in \omega$ if $i \rightarrow j$ for some $j \in \tau$, then $i \rightarrow j$ for every $j \in \tau$ (see Figure 4). Note that there are no constraints on the edges from τ back to ω or on edges within τ or ω .

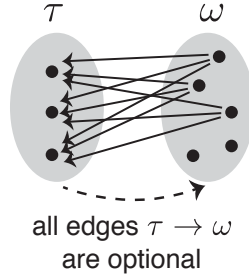


Figure 4: In this graph, ω is simply-added to τ and thus each $i \in \omega$ either sends all possible edges to τ , or no edges. There are no constraints on the edges within τ , within ω , or from τ to ω .

Lemma 2.6 (Lemma 5 in [2]). *Suppose σ has a simply-added split $\tau \dot{\cup} \omega$ with ω simply-added to τ , where τ is uniform in-degree. Let R_τ be the row sum of $I - W_\tau$. Note that this is the maximum (Perron-Frobenius) eigenvalue of $I - W_\tau$. Then*

$$\text{eig}(I - W_\sigma) \supset \text{eig}(I - W_\tau) \setminus R_\tau.$$

So all the eigenvalues of τ get inherited, except possibly the top one R_τ .

As a consequence of Lemma 2.6 and [20, Theorem 1.4], we obtain the following corollary that rules out certain graphs as stable motifs.

Corollary 2.7. *Suppose σ has a simply-added split $\tau \dot{\cup} \omega$ with ω simply-added to τ , where τ is uniform in-degree. If τ is not a stable motif, then σ is not a stable motif.*

2.3. Techniques for ruling out stable motifs

In this section, we highlight and synthesize background results that underlie our arguments ruling out graphs as candidate stable motifs. Corollary 2.7 provides one such result in this direction, whenever the graph has a simply-added split with an unstable uniform in-degree subset. More generally, to rule out a σ as a stable motif, we can either show that σ is not permitted or show that $-I + W_\sigma$ has an eigenvalue with positive real part (equivalently, $I - W_\sigma$ has an eigenvalue with negative real part). *Graphical domination*, first defined in [2], is the primary method for showing that σ is not a permitted motif.

Definition 2.8 ([2]). Let G be a graph on n nodes and $\sigma \subseteq [n]$. For $j, k \in \sigma$, we say that k *graphically dominates* j with respect to σ if the following three conditions all hold:

- (1) if $i \rightarrow j$ then $i \rightarrow k$, for every $i \in \sigma \setminus \{j, k\}$, (2) $j \rightarrow k$, (3) $k \not\rightarrow j$

Theorem 2.9 (Theorem 4 (graphical domination) in [2]). *Let G be a graph on n nodes and $\sigma \subseteq [n]$. Suppose k graphically dominates j with respect to σ for some $j, k \in \sigma$. Then σ is not a permitted motif, and so $\sigma \notin \text{FP}(W)$ for any CTLN W with graph G .*

Another method for ruling out σ as a stable motif is *parity* (Theorem 2.10 below). Throughout the following, to simplify analyses, we will focus on the matrix $I - W_\sigma$; then σ is stable precisely when all the eigenvalues of $I - W_\sigma$ have positive real part. We define the *index* of σ as

$$\text{idx}(\sigma) \stackrel{\text{def}}{=} \text{sgn det}(I - W_\sigma).$$

Since the determinant is the product of the eigenvalues, we see that any stable motif σ must have $\text{idx}(\sigma) = +1$. Thus, any σ with $\text{idx}(\sigma) = -1$ is guaranteed to not be a stable motif.

The following theorem, which is a direct consequence of the Poincaré-Hopf Theorem, shows that there is a detailed balance between the indices of all the fixed points of a CTLN.

Theorem 2.10 (parity [24]). *Let $W = W(G, \varepsilon, \delta)$ be a CTLN. Then*

$$\sum_{\sigma \in \text{FP}(W)} \text{idx}(\sigma) = +1.$$

In particular, the total number of fixed points $|\text{FP}(W)|$ is always odd.

As an immediate consequence of Theorem 2.10, we see that a CTLN has at most 2^{n-1} stable fixed points, since these all have index $+1$ and the maximum size of $\text{FP}(W)$ is $2^n - 1$, the number of choices for a nonempty support σ .

Beyond this general upper bound, Theorem 2.10 can be used to explicitly rule out graphs as stable motifs whenever we know which proper subgraphs are permitted motifs that survive to yield fixed points of the full graph. Specifically, if we know that there are an odd number of proper subgraphs of a graph G that survive as fixed points, then by Theorem 2.10, G cannot have a full-support fixed point and thus G is not a permitted motif. If we additionally know the index of all the subgraphs of a graph G that survive as fixed points, the following conclusions can immediately be drawn from the sum of these indices:

$$\sum_{\sigma \in \text{FP}(W); \sigma \neq [n]} \text{idx}(\sigma) = \begin{cases} +1 & \Rightarrow G \text{ is not a permitted motif} \\ +2 & \Rightarrow G \text{ is permitted, but not stable} \\ 0 & \Rightarrow G \text{ is a permitted motif and stability is unknown} \end{cases}$$

The following lemma from [2] gives another tool for determining a motif's index from that of its proper subgraphs of size one less. In particular, this lemma shows that a CTLN can never contain two stable fixed points whose support differ in only a single neuron.

Lemma 2.11 (Lemma 3 (alternation) in [2]). *Let $W = W(G, \varepsilon, \delta)$ be a CTLN. If $\sigma, \sigma \cup \{k\} \in \text{FP}(W)$, for $k \notin \sigma$, are both fixed point supports, then*

$$\text{idx}(\sigma \cup \{k\}) = -\text{idx}(\sigma).$$

For the previous set of tools, we exploited the fact that the determinant of a matrix equals the product of its eigenvalues. For the remainder of this section, we focus on tools that will utilize the fact that the trace of a matrix is the sum of its eigenvalues. We begin with the following key observation we will exploit throughout.

Observation. If the sum of a subset of eigenvalues of $I - W_\sigma$ is bigger than its trace, $\text{Tr}(I - W_\sigma) = |\sigma|$, then $I - W_\sigma$ must have a negative eigenvalue. In this case, the matrix $-I + W_\sigma$ is unstable, and so σ is not a stable motif. In particular, if the maximum eigenvalue $\lambda_{\max}$ of $I - W_\sigma$ is larger than $|\sigma|$, then σ is not a stable motif.

We next collect some useful results for lower bounding the maximum eigenvalue $\lambda_{\max}$.

Theorem 2.12 (Collatz-Wielandt). *Let A be an $n \times n$ matrix with strictly positive entries. Then the Perron-Frobenius eigenvalue $\lambda_{\max}$ of A is given by*

$$\lambda_{\max} = \max_{x \in \mathbb{R}_{\geq 0}^n} \left(\min_{i \in [n]} \frac{[Ax]_i}{x_i} \right),$$

where the maximum is taken over all non-negative vectors x . In particular,

$$\lambda_{\max} \geq \min_{i \in [n]} \frac{[Ax]_i}{x_i},$$

for any non-negative vector x .

Lemma 2.13. *Let A and B be $n \times n$ matrices with strictly positive entries, and suppose $A \geq B$ entrywise. Then*

$$\lambda_{\max}(A) \geq \lambda_{\max}(B).$$

Proof. Since $A \geq B$, we have $\frac{[Ax]_i}{x_i} \geq \frac{[Bx]_i}{x_i}$ for every non-negative vector x . Thus, by Theorem 2.12 (Collatz-Wielandt), $\lambda_{\max}(A) \geq \lambda_{\max}(B)$. $\square$

Since entrywise comparison of CTLN matrices is fully determined by edges in the underlying graphs, we immediately obtain the following corollary. Note that $E(G)$ denotes the set of edges of a graph G .

Corollary 2.14. *Suppose that G and G' are two graphs on n nodes satisfying $E(G) \subseteq E(G')$, i.e. $i \rightarrow j$ in G implies $i \rightarrow j$ in G' . Let W and W' be CTLNs for graphs G and G' respectively. Then*

$$\lambda_{\max}(I - W) \geq \lambda_{\max}(I - W').$$

Proof. Observe that the off-diagonal entries of W and W' are all $1 - \varepsilon$ and $1 + \delta$. Since $E(G) \subseteq E(G')$, whenever $(I - W)_{ij}$ takes on the smaller value $1 - \varepsilon$ due to $j \rightarrow i$ in G , we have $(I - W')_{ij}$ equals $1 - \varepsilon$ as well. Thus, $I - W \geq I - W'$ entrywise, and so by Lemma 2.13 the result follows. $\square$

3. Degree bounds and proofs of Theorems 1.7 and 1.9

In this section, we prove Theorem 1.7 showing that sparse graphs cannot be stable motifs as well as Theorem 1.9, which guarantees that there is a parameter regime in which the only stable motifs are cliques, and thus the conjecture holds within this regime. Interestingly, these results are both corollaries of a single theorem that connects properties of σ , such as its size or maximum in-degree, to a range of δ values in which σ can never be a stable motif. The key to the proof of this theorem is to show that whenever δ lies in the relevant range, $I - W_\sigma$ has an eigenvalue that is larger than the trace, forcing a negative eigenvalue. Toward that end, we begin with a formula for the Perron-Frobenius eigenvalue $\lambda_{\max}$ of $I - W_\sigma$, when σ is exactly one edge away from a clique. To compute the eigenvalue of such a CTLN, we begin with a lemma identifying an eigenvalue and eigenvector of a highly structured matrix. The proof of this lemma is a straightforward computation, and thus is left to the reader as an exercise.

Lemma 3.1. *Let A be an $n \times n$ matrix of the form*

$$A = \begin{bmatrix} 1 & a & \dots & a & b \\ a & 1 & \dots & a & a \\ \vdots & & \ddots & \vdots & \vdots \\ a & a & \dots & a & 1 \end{bmatrix}$$

for any $a, b \in \mathbb{R}$. Then A has an eigenvector $[c \ 1 \ 1 \ \dots \ 1]^t$ with corresponding eigenvalue λ , where

$$c = 1 + \frac{1}{2} \left[-n + \sqrt{n^2 + 4 \left(\frac{b-a}{a} \right)} \right], \quad \lambda = 1 - a + \frac{a}{2} \left[n + \sqrt{n^2 + 4 \left(\frac{b-a}{a} \right)} \right]$$

Furthermore, if $b \geq a > 0$, then $[c \ 1 \ 1 \ \dots \ 1]^t$ is the Perron-Frobenius eigenvector and $\lambda = \lambda_{\max}$.

As a special case of Lemma 3.1, we obtain the formula for an eigenvalue of a graph that is a clique missing exactly one edge.

Corollary 3.2. *Let σ be a clique missing exactly one edge, $j \not\rightarrow i$. Then $I - W_\sigma$ has Perron-Frobenius eigenvalue*

$$\lambda_{\max} = \varepsilon + \frac{1}{2} (1 - \varepsilon) (|\sigma| + \sqrt{|\sigma|^2 + 4(\delta + \varepsilon)/(1 - \varepsilon)}).$$

Proof. Since σ is a clique missing exactly one edge, without loss of generality suppose the missing edge is from the last node to the first node in σ , i.e. $|\sigma| \not\rightarrow 1$. Then $I - W_\sigma$ has the form of the matrix A in Lemma 3.1 with $a = 1 - \varepsilon$ and $b = 1 + \delta$. Since $b \geq a > 0$, the value of $\lambda_{\max}$ follows directly from that result. $\square$

The following theorem gives parameter regimes in which non-cliques are guaranteed to never be stable motifs.

Theorem 3.3. *Let G be an arbitrary graph and $W = W(G, \varepsilon, \delta)$ a corresponding CTLN. For $\sigma \subseteq [n]$, if σ is not a clique, then σ is not a stable motif whenever the CTLN parameters satisfy*

$$\delta > \frac{\varepsilon}{1-\varepsilon}(|\sigma|^2 - |\sigma| - 1).$$

Moreover, if $G|_\sigma$ has maximum in-degree $d_{\max}^{\text{in}} = d < |\sigma| - 1$, then σ is not a stable motif whenever

$$\delta > \frac{\varepsilon d}{|\sigma| - 1 - d}.$$

Proof. Let σ be any subset that is not a clique. Let G' be a graph of size $|\sigma|$ that is a clique missing exactly one edge, as in Corollary 3.2. Since σ is not a clique, $E(G|_\sigma) \subseteq E(G')$, and so by Lemma 2.13,

$$\lambda_{\max}(I - W_\sigma) \geq \lambda_{\max}(I - W') = \varepsilon + \frac{1}{2}(1 - \varepsilon)(|\sigma| + \sqrt{|\sigma|^2 + 4(\delta + \varepsilon)/(1 - \varepsilon)}).$$

Recall that whenever

$$\lambda_{\max}(I - W_\sigma) > |\sigma| = \text{Tr}(I - W_\sigma),$$

$I - W_\sigma$ must have a negative eigenvalue, forcing $-I + W_\sigma$ to be unstable. Solving for δ , it is straightforward to show that $\varepsilon + \frac{1}{2}(1 - \varepsilon)(|\sigma| + \sqrt{|\sigma|^2 + 4(\delta + \varepsilon)/(1 - \varepsilon)}) > |\sigma|$ whenever

$$\delta > \frac{\varepsilon(|\sigma|^2 - |\sigma| - 1 + (2 - |\sigma|)\varepsilon)}{1 - \varepsilon}.$$

Since σ is not a clique, $|\sigma| \geq 2$, and so it suffices to have $\delta > \frac{\varepsilon}{1-\varepsilon}(|\sigma|^2 - |\sigma| - 1)$ to guarantee the previous inequality holds. Then $-I + W_\sigma$ is unstable, and so σ is not a stable motif.

To prove the second statement, suppose σ has maximum in-degree $d_{\max}^{\text{in}} = d < |\sigma| - 1$. Let G' be a graph of size $|\sigma|$ with uniform in-degree d such that $E(G|_\sigma) \subseteq E(G')$.⁸ To complete the proof, we follow a similar argument as above to ensure that

$$\lambda_{\max}(I - W_\sigma) \geq \lambda_{\max}(I - W') > |\sigma| = \text{Tr}(I - W_\sigma),$$

and thereby force $I - W_\sigma$ to have a negative eigenvalue. In this case, we obtain a tighter bound using the fact that when G' has uniform in-degree d , $I - W'$ has the all-ones vector as the Perron-Frobenius eigenvector and thus, $\lambda_{\max}(I - W')$ equals the row sum:

$$\lambda_{\max}(I - W') = |\sigma| + (|\sigma| - 1 - d)\delta - d\varepsilon.$$

Solving $\lambda_{\max}(I - W') > |\sigma|$ for δ , we see that whenever

$$\delta > \frac{\varepsilon d}{|\sigma| - 1 - d}$$

$I - W_\sigma$ is forced to have a negative eigenvalue, and so σ is not a stable motif. □

⁸Note that it is always possible to find such a G' since there are no constraints on the out-degree, so arbitrary edges can be added to $G|_\sigma$ to appropriately fill up the in-degrees.

From Theorem 3.3, we obtain Theorem 1.7 and Theorem 1.9 as immediate corollaries.

Proof of Theorem 1.7. Let $\sigma \subseteq [n]$ such that $G|_\sigma$ has maximum in-degree $d \stackrel{\text{def}}{=} d_{\max}^{\text{in}} < \frac{|\sigma|}{2}$. Then the bound on δ in the second part of Theorem 3.3 reduces to

$$\frac{\varepsilon d}{|\sigma| - 1 - d} \leq \frac{\varepsilon(|\sigma| - 1)/2}{|\sigma| - 1 - (|\sigma| - 1)/2} = \varepsilon.$$

Within the legal parameter range, we are guaranteed that $\varepsilon < 1$, and $\delta > \frac{\varepsilon}{1-\varepsilon} > \varepsilon$. Thus, δ satisfies the bound in Theorem 3.3 across the full legal range of parameters, and so σ is not a stable motif, meaning it cannot support a stable fixed point. $\square$

Theorem 1.9. When the CTLN parameters satisfy $\delta > \frac{\varepsilon}{1-\varepsilon}(n^2 - n - 1)$, we are guaranteed that $\delta > \frac{\varepsilon}{1-\varepsilon}(|\sigma|^2 - |\sigma| - 1)$ for all $\sigma \subseteq [n]$. Thus, by Theorem 3.3, σ is not a stable motif whenever σ is not a clique. Thus, the only stable motifs in this parameter regime are the cliques, and these produce stable fixed points if and only if they are target-free (Theorem 2.4). $\square$

4. Near-cliques and proof of Theorem 1.8

Theorem 3.3 showed that there is a parameter-regime where non-cliques are provably not stable motifs, and that for sparser graphs we can prove instability for a significantly larger parameter range. In this section, we consider the other extreme of particularly dense graphs and prove instability of those motifs as well. Specifically, we build up the machinery to prove Theorem 1.8 showing that directed cliques and graphs containing a clique of size one less are never stable motifs unless they are themselves cliques. We begin by proving an important fact about the structure of directed cliques.

Lemma 4.1. *Let $\tau \subseteq [n]$ be a directed clique of a graph G . Then there exists a unique $\sigma \subseteq \tau$ such that σ is a target-free clique in $G|_\tau$.*

Proof. First, we show the existence of a target-free clique in $G|_\tau$. Since τ is a directed clique, we can label the neurons $1, \dots, t$, where $t = |\tau|$, such that $i \rightarrow j$ whenever $i < j$. If node t is a sink in $G|_\tau$, then the singleton $\{t\}$ is a target-free clique. Otherwise t has at least one outgoing edge in τ ; let $i_1 \in \tau$ be the vertex with largest index such that $t \rightarrow i_1$, so that $\{i_1, t\}$ is a clique. Either this is a target-free clique, or it has at least one target in τ . In the latter case, let i_2 be the target of $\{i_1, t\}$ with the largest index. Clearly, $i_2 < i_1, t$, and so we also have $i_2 \rightarrow i_1$ and $i_2 \rightarrow t$, making $\{i_1, i_2, t\}$ a clique. If it's not a target-free clique, we can again repeat the previous step and add its target of largest index, yielding a larger clique. Eventually, we obtain a clique $\sigma = \{i_1, \dots, i_k, t\}$ that has no targets in τ , and is thus a target-free clique in $G|_\tau$. (If τ is itself a clique, then $\sigma = \tau$.)

To see that σ is unique, suppose ω is another target-free clique of $G|_\tau$. Then ω must contain the vertex t , otherwise t would be a target of ω by the edge rule defining directed cliques. Observe that all other elements of ω must be less than or equal to i_1 since all nodes in ω receive an edge from t and i_1 is the largest such element by construction. Thus $j \rightarrow i_1$ for all $j \in \omega$, since τ is a directed clique, and so $i_1 \in \omega$ since otherwise ω would have a target.

Similarly, since ω is a clique containing $\{i_1, t\}$, we must have $i_2 \in \omega$, since otherwise i_2 would be a target. Continuing in this manner we see that $\sigma \subseteq \omega$. Furthermore, there cannot be a $j \in \omega \setminus \sigma$, since ω being a clique would force j to be a target of σ , but σ is target-free. Thus $\sigma = \omega$, and so σ is unique. $\square$

Proposition 4.2. *Let G be a directed clique and W be any CTLN with graph G . Then $\text{FP}(W) = \{\sigma\}$, where σ is the unique target-free clique of G . In particular, a directed clique is a stable motif if and only if it is a clique.*

Proof. By Lemma 4.1, G contains a unique target-free clique σ , and so $\sigma \in \text{FP}(W)$ and the corresponding fixed point is stable. To show that σ is the only fixed point support, consider any other subset $\omega \subseteq [n]$. If ω is a clique, then it necessarily has a target, since σ is the unique target-free clique, and thus $\omega \notin \text{FP}(W)$ by Theorem 2.5. Otherwise, ω is a directed clique that is not a clique, under the same ordering of the vertices that showed G was a directed clique. Thus, there exists some pair of vertices with only a unidirectional edge between them. Let $j \in \omega$ be the vertex of largest index such that there exists a $k \in \omega$ with $j \rightarrow k$ but $k \not\rightarrow j$, then choose k to have the largest index among the vertices that do not send edges to j (note by the directed clique property, we necessarily have $k > j$). We will show that k graphically dominates j with respect to ω . Consider $i \in \omega$ such that $i \rightarrow j$. If $i < k$, then $i \rightarrow k$ since ω is a directed clique. If $i > k$, we have $k \rightarrow i$ and so we must have $i \rightarrow k$ as well, since otherwise this would contradict the fact that j was the vertex of largest index that has only a unidirectional edge with some node greater than it. Thus, we have (1) $i \rightarrow j$ implies $i \rightarrow k$ for all $i \in \omega \setminus \{j, k\}$, (2) $j \rightarrow k$, and (3) $k \not\rightarrow j$, and so k graphically dominates j with respect to ω . Thus, $\omega \notin \text{FP}(W)$ by Theorem 2.9, and so $\text{FP}(W) = \{\sigma\}$.

Finally, observe that if G is not itself a clique, i.e. $[n] \neq \sigma$, then $[n] \notin \text{FP}(W)$, and so G is not a permitted motif and thus cannot be a stable motif. $\square$

Proposition 4.2 completes the proof of the first part of Theorem 1.8 by showing that directed cliques are only stable motifs when they are themselves cliques. To prove the second part of the theorem, we first need the following lemma showing alternation of the index of fixed point supports that differ in one node.

Lemma 4.3. *Suppose σ is a permitted motif such that $\sigma = \tau \cup \{k\}$ where τ has uniform in-degree d and the in-degree of k satisfies $d_k^{\text{in}} \leq d$. Then $\text{idx}(\sigma) = -\text{idx}(\tau)$.*

Proof. Let W be any CTLN with graph $G|_\sigma$. By hypothesis, $\sigma = \tau \cup \{k\}$ is permitted, and so $\tau \cup \{k\} \in \text{FP}(W_\sigma)$. Additionally $\tau \in \text{FP}(W_\sigma)$ by Theorem 2.4 (uniform in-degree), since $d_k^{\text{in}} \leq d$. Thus by Lemma 2.11 (alternation), $\text{idx}(\tau \cup \{k\}) = -\text{idx}(\tau)$. $\square$

Recall that a necessary condition for σ to be a stable motif is that σ be a permitted motif with index $+1$. The following proposition shows that if σ contains a clique of size 1 less, then it can only be a stable motif if it is itself a clique.

Proposition 4.4. *If σ contains a clique of size $|\sigma| - 1$ and σ is not a clique, then σ is not a stable motif.*

Proof. Let $\sigma = \tau \cup \{k\}$ where τ is a clique of size $|\sigma| - 1$, so τ has uniform in-degree $|\sigma| - 2$. By Lemma 4.3, if k has in-degree $d_k^{in} \leq |\sigma| - 2$, then whenever σ is permitted it has index $\text{idx}(\sigma) = -\text{idx}(\tau) = -1$. Thus, when $d_k^{in} \leq |\sigma| - 2$, if σ is permitted it must be unstable, and so σ is not a stable motif. On the other hand, when $d_k^{in} = |\sigma| - 1$, so that k receives from every node in τ , there must exist a $j \in \tau$ such that $k \not\rightarrow j$, since σ is not a clique. Then k graphically dominates j with respect to σ because condition (1) and (2) are trivially satisfied since $i \rightarrow k$ for all $i \in \sigma \setminus \{k\}$ and (3) holds by choice of j . Thus, σ is not a permitted motif, and in particular not a stable motif, since it contains a graphical domination relationship. $\square$

We can now complete the proof of Theorem 1.8.

Proof of Theorem 1.8. Proposition 4.2 guarantees that a directed clique is a stable motif if and only if it is a clique, while Proposition 4.4 shows that any graph containing a clique of size 1 less cannot be a stable motif unless it is a clique itself. $\square$

5. Composite graphs that are never stable motifs

In this section, we prove that a variety of *composite graphs* can never be stable motifs, including some well known constructions such as disjoint unions and cyclic unions.

Definition 5.1 (composite graph). Given a set of graphs $G_1, \dots, G_m$, and a graph $\hat{G}$ on m nodes, the *composite graph* with *components* G_i and *skeleton* $\hat{G}$ is the graph G constructed by taking the union of all component graphs, and adding edges between components according to the following rule: if $u \in G_i$ and $v \in G_j$, then $u \rightarrow v$ in G if and only if $i \rightarrow j$ in $\hat{G}$. (See Figure 5.)

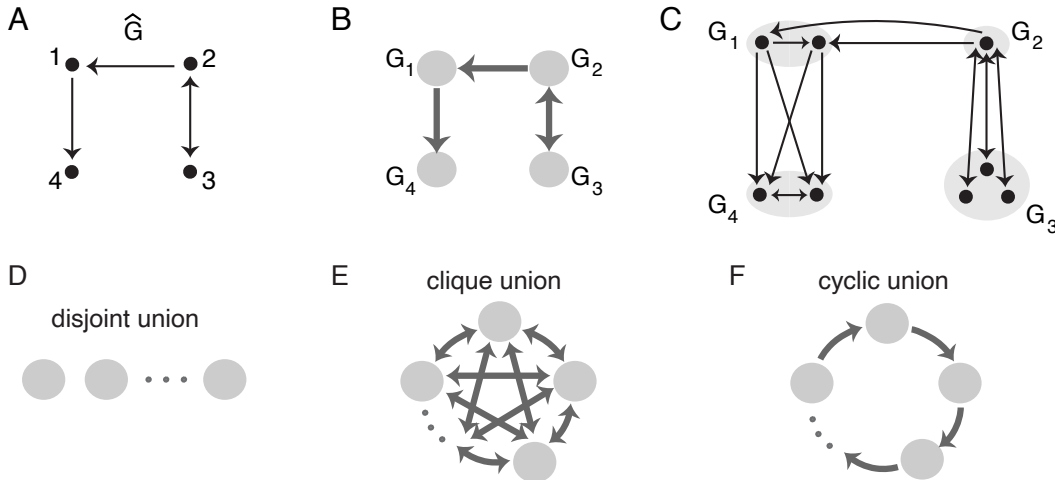


Figure 5: (A) A skeleton graph $\hat{G}$. (B) An arbitrary composite graph with skeleton $\hat{G}$ from A. Each node i in the skeleton is replaced with a component graph G_i whose connections to the rest of the graph are prescribed by the connections of node i in $\hat{G}$. (C) An example composite graph with skeleton $\hat{G}$ from A. (D-F) Families of composite graphs that have previously been studied extensively in [2, 20].

A key feature of composite graphs is that for each component G_i , the rest of the graph treats all the nodes in the component identically in terms of the edges projected to that

component. Specifically, in a composite graph, the rest of the graph is simply-added onto each component, so that $G_i \dot{\cup} (G \setminus G_i)$ is a *simply-added split* for each component (see Section 2.2). This fact was key to proving the following result from [2] that enables us to immediately rule out certain composite graphs as stable motifs.

Theorem 5.2 (Theorem 8 in [2]). *Let G be a composite graph of components $G_1, \dots, G_m$. If some component G_i is not a permitted motif, then G is not a permitted motif. In particular, G is not a stable motif.*

As another consequence of this simply-added split in composite graphs, we can apply Lemma 2.6 whenever one of the components has uniform in-degree. Recall that σ is *uniform in-degree* if all nodes in $G|_\sigma$ have the same in-degree. In this case, all the eigenvalues of the uniform in-degree component are inherited by the full composite graph, except for possibly the Perron-Frobenius eigenvalue $\lambda_{\max}$ of the component. Thus, if a component is an unstable uniform in-degree, then the full composite graph will also be unstable, as captured by the following proposition from [2].

Proposition 5.3 (Proposition 1 in [2]). *Let G be a composite graph with components $G_1, \dots, G_m$. If some component G_i is an unstable uniform in-degree motif, then G is not a stable motif.*

These results enable us to rule out as possible stable motifs all composite graphs where either a component is not permitted or it is an unstable uniform in-degree. But what about the case when all the components are permitted and stable, specifically, what if all the components are cliques? The following theorem shows that as long as the composite graph satisfies a particular bound on the in-degree of the nodes, it will not be a stable motif even when all the components are stable motif cliques.

Theorem 5.4. *Let G be a composite graph on n nodes with skeleton $\hat{G}$ on m nodes, and suppose the components $G_1, \dots, G_m$ are all cliques (of arbitrary size). Let $d_{\max}^{\text{in}}(G)$ be the maximum in-degree of all nodes in G . If*

$$d_{\max}^{\text{in}}(G) \leq n - \frac{m+1}{2},$$

then G is not a stable motif. In particular, if the skeleton $\hat{G}$ has maximum in-degree $d_{\max}^{\text{in}}(\hat{G}) < \frac{m}{2}$, then G is not a stable motif.

Proof. Let the components $G_1, \dots, G_m$ be cliques of sizes $n_1, \dots, n_m$ respectively so that $n = n_1 + \dots + n_m$. Recall that a clique σ has $\lambda_{\max} = |\sigma|(1 - \varepsilon) + \varepsilon$, while the remaining $|\sigma| - 1$ eigenvalues all equal ε (see Theorem 2.5). Since the rest of the graph is simply-added onto each component, Lemma 2.6 guarantees that $I - W$ inherits $n_i - 1$ eigenvalues, all equal to ε , from each component G_i . Hence, ε is an eigenvalue of $I - W$ with multiplicity $\sum_{i=1}^m (n_i - 1) = n - m$.

We will show that the sum of $\lambda_{\max}(I - W)$ with these $n - m$ eigenvalues ε is strictly greater than $\text{Tr}(I - W) = n$, guaranteeing that $I - W$ has a negative eigenvalue and G is not a stable motif. Recall that by Theorem 2.12 (Collatz-Wielandt), $\lambda_{\max} \geq \min_{i \in [n]} \frac{[(I - W)x]_i}{x_i}$, for every nonnegative vector x . We will use this to show that $\lambda_{\max} > n - (n - m)\varepsilon$. Observe that when

x is the all-ones vector, $\frac{[(I-W)x]_i}{x_i}$ is simply the i^{th} row sum R_i of $I - W$, which is determined by the number of inputs d_i^{in} to node i in G . Thus, for $x = [1 \dots 1]^t$, we have

$$\frac{[(I-W)x]_i}{x_i} = R_i = n - \varepsilon d_i^{\text{in}} - \delta(n - 1 - d_i^{\text{in}}).$$

Next observe that $R_i > n - (n - m)\varepsilon$ precisely when

$$\varepsilon(n - m - d_i^{\text{in}}) + \delta(n - 1 - d_i^{\text{in}}) > 0. \quad (4)$$

Solving for d_i^{in} , we see that this inequality is satisfied when $d_i^{\text{in}} < \frac{\varepsilon(n-m) + \delta(n-1)}{\varepsilon + \delta}$.

Next since

$$\begin{aligned} \frac{\varepsilon(n - m) + \delta(n - 1)}{\varepsilon + \delta} &= \frac{(\varepsilon + \delta)(n - m) + \delta(m - 1)}{\varepsilon + \delta} \\ &= n - m + \frac{\delta(m - 1)}{\varepsilon + \delta} \\ &> n - m + \frac{\delta(m - 1)}{2\delta} \\ &= n - \frac{m + 1}{2}, \end{aligned}$$

we see that whenever $d_i^{\text{in}} \leq n - \frac{m+1}{2}$, the inequality (4) is satisfied. Thus, by Theorem 2.12 (Collatz-Wielandt), $\lambda_{\max} > n - (n - m)\varepsilon$, and so $I - W$ must have a negative eigenvalue. Therefore, G is not a stable motif.

To prove the last statement of the theorem, suppose that the skeleton $\hat{G}$ has maximum in-degree $d_{\max}^{\text{in}}(\hat{G}) < \frac{m}{2}$, so that $\hat{d}_j^{\text{in}} \leq \frac{m-1}{2}$ for every node j in the skeleton graph. Then in the composite graph G , for every node in component G_j , there are at least $m - \frac{m-1}{2} = \frac{m+1}{2}$ components that the node does not receive from in G . Since every component is nonempty, this guarantees that each node i in G_j has at least $\frac{m+1}{2}$ nodes in G that it does not receive from, and so $d_i^{\text{in}} \leq n - \frac{m+1}{2}$. Applying the first part of the theorem we then see that G is not a stable motif. $\square$

As an immediate corollary, we see that when the skeleton $\hat{G}$ is an independent set or a cycle, so that G is a disjoint union or cyclic union respectively, then G is not a stable motif.

Corollary 5.5. *Let G be a disjoint union or a cyclic union of cliques. Then G is not a stable motif.*

6. Stable motifs up to size 4 and proof of Theorem 1.6

In this section, we prove Theorem 1.6 showing that for σ of size at most 4, σ is a stable motif if and only if it is a clique, and thus it can only support a stable fixed point in a larger network if it is a target-free clique. For the proof, we analyze all permitted motifs up through size 4 and show that each non-clique cannot be a stable motif by applying key results from each of the earlier sections. To perform this analysis, we must first generate a comprehensive list of these permitted motifs; the following lemma is key to efficiently generating this list.

Lemma 6.1. *Let σ have size $|\sigma| \leq 4$ and suppose σ is a permitted motif of a CTLN $W = W(G, \varepsilon, \delta)$ for some legal choice of ε and δ . Then σ is a permitted motif for all ε and δ in the legal range. Furthermore, $\text{idx}(\sigma)$ is constant across all ε and δ in the legal range.*

Proof. In [2, Theorem 6], it was shown that the collection of fixed point supports, $\text{FP}(W)$, is parameter independent for all graphs up to size 4. Thus when $|\sigma| \leq 4$, if σ is a permitted motif for a particular choice of ε and δ , then σ is a permitted motif for all ε and δ in the legal range.

Furthermore, the proof of [2, Theorem 6] established that all permitted motifs up to size 3 have parameter-independent survival rules and have index that is constant across all ε and δ in the legal range. Since the index of a permitted motif of size 4 can be determined in a parameter-independent way via Theorem 2.10 (parity) from the sum of the indices of the surviving permitted motifs contained in it, it follows that $\text{idx}(\sigma)$ is constant across all ε and δ in the legal range for σ of size 4 as well. $\square$

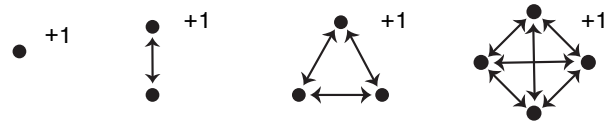
It was shown in [2] that there are permitted motifs of size 4 that have parameter-dependent survival in larger networks, and thus there are graphs of size 5 that are only permitted motifs for ε and δ within a subset of the legal range (see Appendix A.2 in [2]). Thus, Lemma 6.1 does not extend to larger n .

Figure 6 shows all the permitted motifs of size $|\sigma| \leq 4$ together with their index. These were identified computationally by finding all graphs that have a full-support fixed point for $\varepsilon = 0.25$ and $\delta = 0.5$, which Lemma 6.1 guarantees is sufficient to generate a complete, parameter-independent list. It is worth noting that while all the indices were determined computationally, many of them could have actually been determined directly via Lemma 4.3 since they contain a uniform in-degree subgraph of size 1 less.

As an aside, one interesting observation from Figure 6 is that whenever a graph is permitted, its complement is also permitted. Furthermore, for graphs of size 3, the index of a graph and its complement are the same, whereas for size 4, complementing flips the sign of the index. It turns out that the pattern of a graph being permitted if and only if its complement is permitted does not hold for size 5 however (this breaks down in many of the cases when the graph contains a subgraph of size 4 that has parameter-dependent survival).

We now turn to the proof of Theorem 1.6, in which we will show that the only graphs in Figure 6 that are stable motifs are precisely the cliques (shown in Panel A). To aid the proof, the graphs are organized in the figure by size and then by which results will be used to show they are not stable motifs.

A $|\sigma| \leq 4$ stable motifs

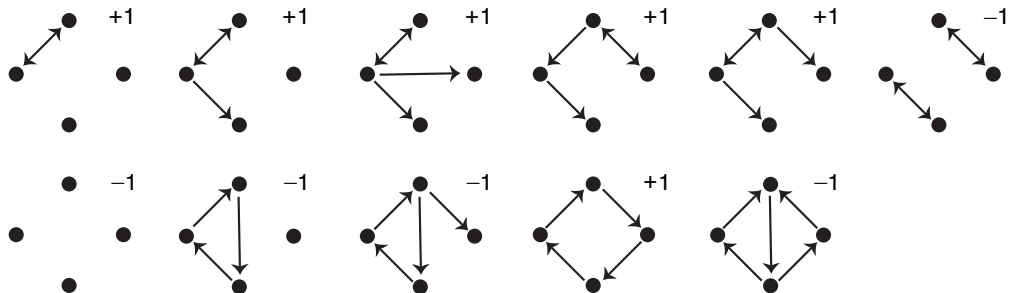


$|\sigma| \leq 4$ permitted motifs that are not cliques

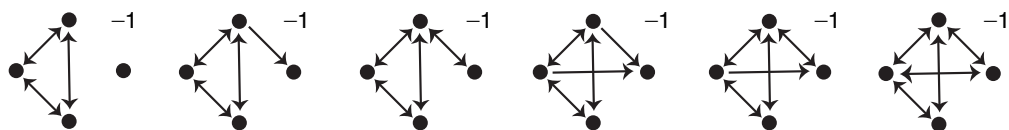
B



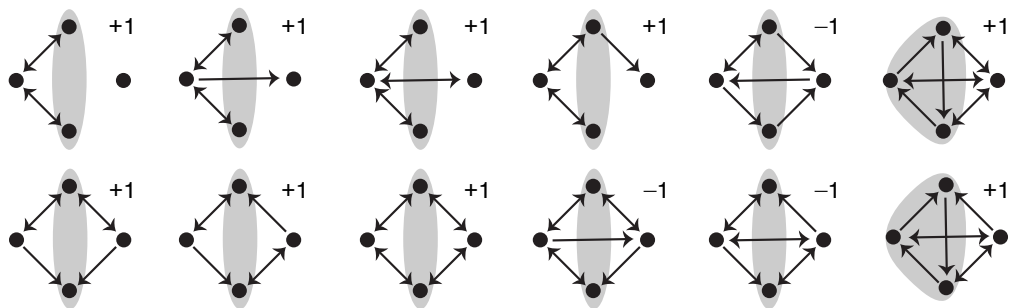
C



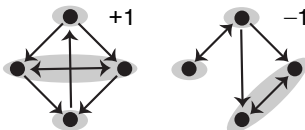
D contains clique of size $|\sigma| - 1$



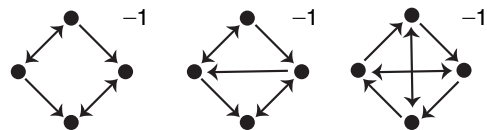
E simply-added onto unstable uniform in-degree



F composite graph with sparse skeleton



G miscellaneous $\text{idx}(\sigma) = -1$



H Collatz-Weilandt argument

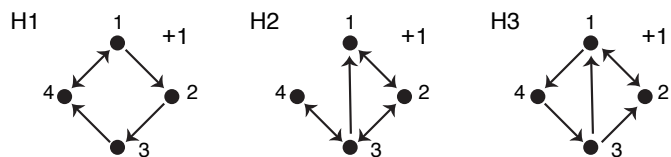


Figure 6: All permitted motifs with $|\sigma| \leq 4$ annotated with their index on the top right of the graph.

Proof of Theorem 1.6. Panel B of Figure 6 shows all the permitted motifs of size $|\sigma| \leq 3$ that are not cliques. Observe that all these graphs have maximum in-degree $d_{\max}^{\text{in}} \leq 1 < \frac{|\sigma|}{2}$ and/or contain a clique of size $|\sigma| - 1 = 2$. Thus, by Theorems 1.7 and 1.8, none of these graphs is a stable motif.

Next consider the permitted motifs of size 4 that are not cliques. All the graphs in panel C of Figure 6 have maximum in-degree $d_{\max}^{\text{in}} \leq 1 < \frac{|\sigma|}{2}$ and/or are oriented, and so by Theorems 1.7 and 1.4 are not stable motifs. The graphs in panel D all contain a clique of size $|\sigma| - 1 = 3$, and thus by Theorem 1.8 are not stable motifs.

For panel E of Figure 6, observe that every graph has a simply-added split $\sigma = \tau \dot{\cup} \omega$ with ω simply-added onto a τ (the nodes in the gray shaded region) that is uniform in-degree. Since τ is either an independent set or a 3-cycle, and neither of these is a stable motif, Corollary 2.7 guarantees that σ also is not a stable motif.

The two graphs in panel F can be decomposed as composite graphs where each component is a clique. The skeleton of the first composite graph is a 3-cycle, while the skeleton of the second graph is a 2-clique with an outgoing edge. Both these skeletons have maximum in-degree 1, and thus by Theorem 5.4, neither composite graph is a stable motif.

The graphs in panel G do not fit into any of the families characterized by earlier results, but they all have index -1 , and thus cannot be stable motifs.

Finally, for the graphs in panel H of Figure 6, we must appeal to more detailed arguments about the eigenvalues to show they are not stable motifs. Specifically, we will use Theorem 2.12 (Collatz-Wielandt) to show that these graphs have a maximum eigenvalue (or sum of eigenvalues) that is larger than the trace, thus forcing $I - W_\sigma$ to have a negative eigenvalue, and guaranteeing that σ is not a stable motif.

(H1) For the graph in H1,

$$I - W_\sigma = \begin{bmatrix} 1 & 1 + \delta & 1 + \delta & 1 - \varepsilon \\ 1 - \varepsilon & 1 & 1 + \delta & 1 + \delta \\ 1 + \delta & 1 - \varepsilon & 1 & 1 + \delta \\ 1 - \varepsilon & 1 + \delta & 1 - \varepsilon & 1 \end{bmatrix}.$$

Consider $x = [1, 1, 1, 1 - \delta/2]^T$, then

$$[(I - W_\sigma)x]_1 = 3 + 2\delta + 1 - \varepsilon - \delta/2 + \varepsilon\delta/2 > 4.$$

$$[(I - W_\sigma)x]_2 = [(I - W_\sigma)x]_3 = 4 + \delta/2 - \delta^2/2 + \delta - \varepsilon > 4.$$

$$[(I - W_\sigma)x]_4 = 4 + \delta/2 - 2\varepsilon > 4 - 2\varepsilon > 4(1 - \delta/2).$$

Thus, by Theorem 2.12 (Collatz-Wielandt), $\lambda_{\max} \geq \min_{i \in \sigma} \frac{[(I - W_\sigma)x]_i}{x_i} > 4$. Since $\lambda_{\max}$ exceeds $\text{Tr}(I - W_\sigma) = 4$, σ is not a stable motif.

(H2) First observe that this graph has a simply-added split where $\{3, 4\}$ is simply-added onto the clique $\{1, 2\}$. Thus, by Lemma 2.6, σ inherits the eigenvalue ε from the clique. Therefore, we need only show that $\lambda_{\max}$ of $I - W_\sigma$ exceeds $4 - \varepsilon$ to guarantee that a

sum of eigenvalues exceeds $\text{Tr}(I - W_\sigma) = 4$. Observe that

$$I - W_\sigma = \begin{bmatrix} 1 & 1 - \varepsilon & 1 - \varepsilon & 1 + \delta \\ 1 - \varepsilon & 1 & 1 - \varepsilon & 1 + \delta \\ 1 + \delta & 1 - \varepsilon & 1 & 1 - \varepsilon \\ 1 + \delta & 1 + \delta & 1 - \varepsilon & 1 \end{bmatrix}$$

and consider $x = [1, 1, 1, 1 + \varepsilon/2]^T$. Then

$$[(I - W_\sigma)x]_1 = [(I - W_\sigma)x]_2 = 3 - 2\varepsilon + 1 + \delta + (1 + \delta)\varepsilon/2 > 4 - 2\varepsilon + \delta > 4 - \varepsilon.$$

$$[(I - W_\sigma)x]_3 = 3 + \delta - \varepsilon + 1 - \varepsilon + (1 - \varepsilon)\varepsilon/2 > 4 - 2\varepsilon + \delta > 4 - \varepsilon.$$

$$[(I - W_\sigma)x]_4 = 4 + 2\delta - \varepsilon/2 > 4 + \varepsilon > 4 + \varepsilon - \varepsilon^2/2 = (4 - \varepsilon)(1 + \varepsilon/2).$$

Thus by the Collatz-Wielandt Theorem, $\lambda_{\max} > 4 - \varepsilon$, and $\lambda_{\max} + \varepsilon > \text{Tr}(I - W_\sigma)$.

Therefore, σ is not a stable motif.

(H3) As with the previous graph, there is a simply-added split where $\{3, 4\}$ is simply-added onto the clique $\{1, 2\}$, and so σ inherits the eigenvalue ε from the clique. We will again use Theorem 2.12 (Collatz-Wielandt) to show that $\lambda_{\max} > 4 - \varepsilon$. Observe that

$$I - W_\sigma = \begin{bmatrix} 1 & 1 - \varepsilon & 1 - \varepsilon & 1 + \delta \\ 1 - \varepsilon & 1 & 1 - \varepsilon & 1 + \delta \\ 1 + \delta & 1 + \delta & 1 & 1 - \varepsilon \\ 1 - \varepsilon & 1 + \delta & 1 + \delta & 1 \end{bmatrix}$$

and again consider $x = [1, 1, 1, 1 + \varepsilon/2]^T$. Then

$$[(I - W_\sigma)x]_1 = [(I - W_\sigma)x]_2 = 3 - 2\varepsilon + 1 + \delta + (1 + \delta)\varepsilon/2 > 4 - 2\varepsilon + \delta > 4 - \varepsilon.$$

$$[(I - W_\sigma)x]_3 = 3 + 2\delta + 1 - \varepsilon + (1 - \varepsilon)\varepsilon/2 = 4 - \varepsilon + 2\delta + (1 - \varepsilon)\varepsilon/2 > 4 - \varepsilon.$$

$$[(I - W_\sigma)x]_4 = 4 + 2\delta - \varepsilon/2 > 4 + \varepsilon > 4 + \varepsilon - \varepsilon^2/2 = (4 - \varepsilon)(1 + \varepsilon/2).$$

Thus, $\lambda_{\max} > 4 - \varepsilon$, and so $I - W_\sigma$ has a sum of eigenvalues larger than its trace, forcing a negative eigenvalue. Hence σ is not a stable motif.

Therefore every permitted motif up to size 4 that is not a clique is unstable, and so if $|\sigma| \leq 4$, then σ supports a stable fixed point if and only if σ is a target-free clique. $\square$

7. Enumerating target-free cliques and proof of Theorem 1.10

Thus far we have seen that for certain parameter regimes, the only stable motifs are cliques, and we have also ruled out a large variety of graphs from being stable motifs for every choice of parameters within the legal range. This provides strong evidence for the conjecture that the only stable motifs are cliques. Furthermore, since cliques only survive to yield fixed points of the full network when they are target-free, a critical step to understanding the collection of stable fixed points of a CTLN is to be able to find and count the target-free cliques of its underlying directed graph.

In this section, we prove a correspondence between target-free cliques of a directed graph and maximal cliques of an undirected graph. Counting and enumerating maximal

cliques of undirected graphs has garnered significant attention in extremal graph theory, and thus this correspondence will enable us to easily import results from this field to yield information about the target-free cliques of a directed graph, and thus about the stable fixed points of corresponding CTLNs.

Given a directed graph G , we can construct a corresponding undirected graph $\tilde{G}$ by making all bidirectional edges in G into undirected edges in $\tilde{G}$ and dropping all other edges, i.e. (i, j) is an undirected edge in $\tilde{G}$ if and only if $i \leftrightarrow j$ in G . Then every target-free clique of G corresponds to a maximal clique in $\tilde{G}$ (although $\tilde{G}$ may have additional maximal cliques as well that were originally targeted in G). Thus, the number of target-free cliques in G is at most the number of maximal cliques in $\tilde{G}$. In fact, the following lemma shows that the maximum number of target-free cliques in directed graphs actually equals the maximum number of maximal cliques in undirected graphs.

Lemma 7.1. *The maximum number of target-free cliques in a directed graph equals the maximum number of maximal cliques in an undirected graph.*

Proof. From the above construction, we see that the maximum number of target-free cliques in a directed graph is less than or equal to the maximum number of maximal cliques in an undirected graph. To see the reverse inequality, observe that given any undirected graph, there is a canonical corresponding directed graph obtained by replacing each undirected edge with a bidirectional edge. In this case, all the maximal cliques of the undirected graph yield target-free cliques in the corresponding directed graph. Thus, the maximum number of target-free cliques in a directed graph is greater than or equal to the maximum number of maximal cliques in an undirected graph, and so equality holds. $\square$

Lemma 7.1 allows us to immediately translate upper bounds on the number of maximal cliques in an undirected graph into upper bounds on the number of target-free cliques in directed graphs, and thus into upper bounds on the number of stable fixed points in CTLNs in parameter regimes where the conjecture holds. For example, applying an upper bound from Moon and Moser [23] immediately yields Theorem 1.10, restated below.

Theorem 1.10. *The maximum number of target-free cliques in a directed graph of size n is*

$$\max \# \text{ of target-free cliques} = \begin{cases} 3^{n/3} & \text{if } n \equiv 0 \pmod{3} \\ 4 \cdot 3^{\lfloor n/3 \rfloor - 1} & \text{if } n \equiv 1 \pmod{3} \\ 2 \cdot 3^{\lfloor n/3 \rfloor} & \text{if } n \equiv 2 \pmod{3}. \end{cases}$$

Lemma 7.1 allows us to import a number of other results from the extremal graph theory literature as well. For example, Hedman gives a tighter upper bound on the number of maximal cliques whenever the size of the largest clique is at most $n/2$ [25]. Moon and Moser also give an upper bound on the number of different sizes of maximal cliques [23], while Spencer gives a lower bound on this for particularly large n [26], which improves on a previous lower bound of Erdős [27].

Additionally, the construction of an undirected $\tilde{G}$ from a directed graph G enables us to apply algorithms for finding maximal cliques in order to enumerate all the target-free cliques

of G . Specifically, the list of maximal cliques of $\tilde{G}$ gives all the maximal cliques of G , and it is straightforward to check which of these candidate cliques are in fact target-free in G . Thus, target-free cliques can be easily found using algorithms for finding maximal cliques such as those from Bron and Kerbosch [28] (which has worst case running time of $O(3^{n/3})$), based on a branch-and-bound technique, or that of Tomita et al. [29], based on a depth-first search algorithm with pruning.

Acknowledgments. All three authors were supported by NIH R01 EB022862. The first author was also supported by NSF DMS-1516881 and NSF DMS-1951165, while the third author was also supported by NSF DMS-1951599.

References

- [1] K. Morrison, A. Degeratu, V. Itskov, and C. Curto. Diversity of emergent dynamics in competitive threshold-linear networks: a preliminary report. Available at <https://arxiv.org/abs/1605.04463>
- [2] C. Curto, J. Geneson, and K. Morrison. Fixed points of competitive threshold-linear networks. *Neural Comput.*, 31(1):94–155, 2019.
- [3] J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci.*, 79(8):2554–2558, 1982.
- [4] Daniel J. Amit. *Modeling brain function: The world of attractor neural networks*. Cambridge University Press, 1989.
- [5] X. Xie, R. H. Hahnloser, and H.S. Seung. Selectively grouping neurons in recurrent networks of lateral inhibition. *Neural Comput.*, 14:2627–2646, 2002.
- [6] C. Curto and K. Morrison. Pattern completion in symmetric threshold-linear networks. *Neural Computation*, 28:2825–2852, 2016.
- [7] C. Hillar, R. Mehta, and K. Koepsell. A hopfield recurrent neural network trained on natural images performs state-of-the-art image compression. *2014 IEEE International Conference on Image Processing (ICIP), Paris, France*, pages 4092–4096, 2014.
- [8] D. Ferster and K. D. Miller. Neural mechanisms of orientation selectivity in the visual cortex. *Annual review of neuroscience*, 2000.
- [9] H. Ozeki, I.M. Finn, E.S. Schaffer, K.D. Miller, and D. Ferster. Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron*, 62(4):578–592, 2009.
- [10] M. Tsodyks. Attractor neural network models of spatial maps in hippocampus. *Hippocampus*, 9:481–489, 1999.
- [11] R. Moreno-Bote, J. Rinzel, and N. Rubin. Noise-induced alternations in an attractor network model of perceptual bistability. *J. Neurophysiol.*, 98(3):1125–1139, 2007.

- [12] A. Piet, J. C. Erlich, and C. D. Brody. Rat prefrontal cortex inactivations during decision making are explained by bistable attractor dynamics. *Neural Comput.*, 29(11):2861–2886, 2017.
- [13] H.S. Seung and R. Yuste. *Principles of Neural Science*, chapter Appendix E: Neural networks, pages 1581–1600. McGraw-Hill Education/Medical, 5th edition, 2012.
- [14] M. Simmen, A. Treves, and E. Rolls. Pattern retrieval in threshold-linear associative nets. *Network: Computation in Neural Systems*, 7:109–122, 1996.
- [15] R. H. Hahnloser, R. Sarpeshkar, M.A. Mahowald, R.J. Douglas, and H.S. Seung. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature*, 405:947–951, 2000.
- [16] R. H. Hahnloser, H.S. Seung, and J.J. Slotine. Permitted and forbidden sets in symmetric threshold-linear networks. *Neural Comput.*, 15(3):621–638, 2003.
- [17] C. Curto, A. Degeratu, and V. Itskov. Flexible memory networks. *Bull. Math. Biol.*, 74(3):590–614, 2012.
- [18] C. Curto, A. Degeratu, and V. Itskov. Encoding binary neural codes in networks of threshold-linear neurons. *Neural Comput.*, 25:2858–2903, 2013.
- [19] T. Biswas and J. E. Fitzgerald. A geometric framework to predict structure from function in neural networks. Available at <https://arxiv.org/abs/2010.09660>
- [20] C. Parmelee, J. Londono Alvarez, C. Curto, and K. Morrison. Sequential attractors in combinatorial threshold-linear networks. *SIAM J. Appl. Dyn. Syst.*, 21(2), 2022. Available at <https://arxiv.org/abs/2107.10244>
- [21] D. Egas Santander, S. Ebli, A. Patania, N. Sanderson, F. Burtscher, K. Morrison, and C. Curto. *Research in Computational Topology 2*, chapter Nerve theorems for fixed points of neural networks. Assoc. Women Math. Ser. 30. Springer, 2022.
- [22] C. Parmelee, S. Moore, K. Morrison, and C. Curto. Core motifs predict dynamic attractors in combinatorial threshold-linear networks. *PLOS ONE*, 17, 2022.
- [23] J. W. Moon and L. Moser. On cliques in graphs. *Israel J. Math.*, 3:23–28, 1965.
- [24] K. Morrison, J. Geneson, C. Langdon, A. Degeratu, V. Itskov, and C. Curto. Emergent dynamics from network connectivity: a minimal model. *In preparation*.
- [25] B. Hedman. The maximum number of cliques in dense graphs. *Discrete Mathematics*, 54:161–166, 1985.
- [26] J. Spencer. On cliques in graphs. *Israel J. Math.*, 9:419–421, 1971.
- [27] P. Erdős. On cliques in graphs. *Israel J. Math.*, 4:233–234, 1966.

- [28] C. Bron and J. Kerbosch. Algorithm 457: Finding all cliques of an undirected graph. *Comm. ACM*, 16:48–50, 1973.
- [29] E. Tomita, A. Tanaka, and H. Takahashi. The worst-case time complexity for generating all maximal cliques and computational experiments. *Theor. Comput. Sci.*, 363:28–42, 2006.