Instability of Graph Dynamical Systems and Edge Shattering

Matthew Macauley

Department of Mathematical Sciences
Clemson University
Clemson, South Carolina, USA 29634

Special session: Biomathematics: Newly developed applied mathematics and new mathematics arising from biosciences
AMS-SMS Joint Meeting
Fudan University
Shanghai, China
December, 2008

Outline

Definitions

- A *graph dynamical system* (GDS) is a triple consisting of:
  - A graph Y with vertex set $\mathrm{v}[Y] = \{1, 2, \ldots, n\}$.
  - For each vertex i a state $y_i \in K$ (e.g. $\mathbb{F}_2 = \{0, 1\}$) and a Y-local function $F_i \colon K^n \longrightarrow K^n$

$$F_i(\mathbf{y} = (y_1, y_2, \ldots, y_n)) = (y_1, \ldots, y_{i-1}, \underbrace{f_i(\mathbf{y}[i])}_{\text{vertex function}}, y_{i+1}, \ldots, y_n)\ .$$

  - An update scheme that determines how to assemble the functions to obtain the *global map* $\mathbf{F} \colon K^n \longrightarrow K^n$.

- Two standard choices for the update scheme:
  - Parallel: *Generalized cellular automata*

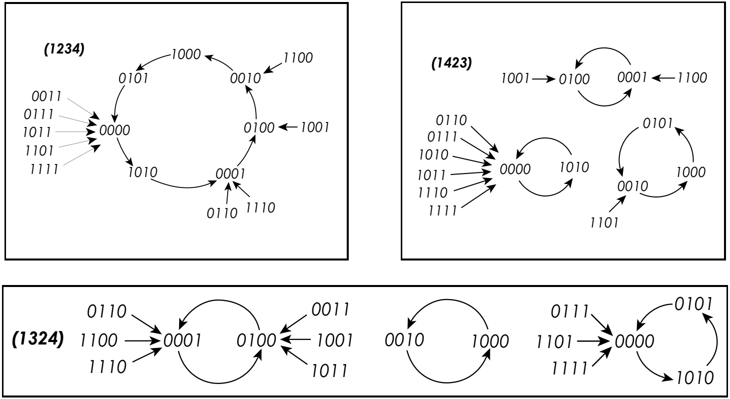
$$\boxed{\mathbf{F}(x_1, \ldots, x_n) = (f_1(x[1]), \ldots, f_n(x[n])).}$$

  - Sequential: *Sequential dynamical system*

$$\boxed{[\mathfrak{F}_Y, \pi] = F_{\pi_n} \circ F_{\pi_{n-1}} \circ \cdots \circ F_{\pi_1}\ .}$$

Example. Define the function $\mathrm{nor}_k \colon \mathbb{F}_2^k \longrightarrow \mathbb{F}_2$ by $\mathrm{nor}_k(\mathbf{x}) = \prod_{i=1}^{k}(1 + x_i)$.

- $[\mathbf{Nor}_{\mathrm{Circ}_4}, \pi]$ for given update sequences:



Applications

- Large complex networks.
  - Epidemiology. Disease propogation over social contact graphs.
  - Agent-based transportation simulations.
  - Packet flow in wireless networks.
- Gene annotation (functional linkage networks)
- Transport computation on irregular grids (e.g., heat, radiation).
- Image processing and pattern recognition.
- Discrete event simulations (e.g., chemical reaction networks).

Generic GDS research questions

- What does it mean for two GDSs to be "equivalent"?
- What is a good way to measure "stability" of the dynamics of a GDS with respect to changes in the system (e.g., functions, states, update sequence)?
- How is update stability correlated with properties of the system, such as the structure of the base graph?
- *What is a good characterization of graphs that is useful to people studying dynamical systems over them?*

Functional Linkage Networks

- Begin with a *functional-linkage graph*:
  - Corresponds to a Gene Ontology (GO) function, f.
  - Vertices of a graph represent proteins.
  - Two proteins are adjacent if we think they share the same function. The edge weight w_{ij} is our level of certainty.
  - Each protein is assigned a state x_i from $\{+1, -1, 0\}$, depending on whether it is annotated with f. ($+1 =$ yes, $-1 =$ no, $0 =$ mystery).

Goal: Assign a value of $+1$ or -1 to all "mystery proteins."

Basic approach: Given "mystery protein" i, is it adjacent to more $+1$ or more -1 proteins? i.e., compute

$$s_i = \mathrm{sign}\left(\sum_{j\,|\,\{i,j\}\in E} w_{ij}x_j - \theta\right) .$$

- If $s_i > \theta$, then assign $x_i = +1$. Otherwise, set $x_i = -1$.

Functional Linkage Networks (cont.)

In [2], the proteins are updated sequentially, given by an update order chosen at random. This is an SDS.

Some general questions:

- Does this process always converge to a fixed point? (Answer: YES)
- How much does the fixed point reached depend on the update order used?
- How quickly does it take to reach a fixed point?
- How reliable is this algorithm? (i.e., false negative & false positive rate)

In fact, this method has been well-received, and the reliability is superior to prior models. But as with any model, there are some “red flags” that are worth investigating.

A Mathematical Abstraction with 2-Threshold SDSs

Motivation: Let's explore one of the "red flags" of this approach.

- Given $\mathfrak{F}_Y$, define
$$\omega_\pi(\mathbf{y}) = \bigcap_{n=1}^{\infty} \{[\mathfrak{F}_Y, \pi]^m(\mathbf{y}) \mid m \geq n\} \ .$$
- For $\mathcal{P} \subseteq S_Y$, define
$$\omega_{\mathcal{P}}(\mathbf{y}) = \bigcup_{\pi \in \mathcal{P}} \omega_\pi(\mathbf{y}) \ .$$
- For a sequence of functions $\mathfrak{F}_Y$, define
$$\omega(\mathfrak{F}_Y) = \max \left\{ \left| \omega_{S_Y}(\mathbf{y}) \right| \mid \mathbf{y} \in K^n \right\} \ .$$
- Consider an SDS $[\mathfrak{F}_Y, \pi]$, where $\mathfrak{F}_Y = \mathsf{T}^2_Y$, the 2-threshold local functions.
- All periodic points of a threshold SDS are fixed points.

Summary of Results

- A threshold SDS over K_n can have at most $n+1$ fixed points, and this bound is sharp.
- If $Y = K_n$ and $k \leq n$, then $\mathrm{Fix}[\mathrm{T}^k_Y, \pi] = \{\mathbf{0}, \mathbf{1}\}$.
- If Y is connected with minimial degree $d > (1 - \frac{1}{k})n$, for $k > 0$, then $\mathrm{Fix}[\mathrm{T}^k_Y, \pi] \subseteq \{\mathbf{0}, \mathbf{1}\}$.
- If Y is a tree, then $\Omega(2^n)$ fixed points can be reached by varying the update order of an SDS $[\mathrm{T}^2_Y, \pi]$.

Theorem ([4])

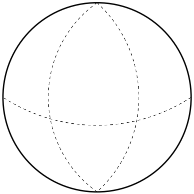
Let $0 < \epsilon < 1$. Threshold systems over $G_{n,p}$, with $p = o\left(\frac{n^\epsilon}{n}\right)$, contain initial configurations from which $\Omega(2^{n^{1-\epsilon}})$ different fixed points can be reached by changing the update order, with probability $1 - o\left(\frac{1}{n^\epsilon}\right)$

Conclusion. *In general, adding edges to the dependency graph of an SDS causes the dynamics to become more stable with respect to update order.*

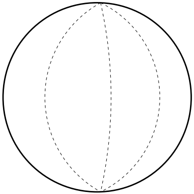
Hyperplane arrangements – a motivating example

Consider the problem of counting the number of chambers $|\mathcal{C}(\mathcal{H})|$ of a hyperplane arrangement $\mathcal{H}$ in $\mathbb{R}^n$.

Example:



3 cuts of S^2
8 chambers



3 cuts of S^2
6 chambers

Figure: Cutting the sphere with hyperplanes

The number of chambers depends not only on the number of hyperplanes, but also on the linear dependencies of the normal vectors. This is a problem handled by *matroids*.

Consider a hyperplane arrangement $\mathcal{H} = \{H_1, \ldots, H_k\}$, with corresponding normal vectors $\mathcal{V} = \{v_1, \ldots, v_k\}$.

If the normal vectors are linearly independent, then $|\mathcal{C}(\mathcal{H})| = 2^k$.

If the hyperplanes (normal vectors) are in *general position*, $|\mathcal{C}(\mathcal{H})| = 2\sum_{i=0}^{n-1}\binom{k-i}{i}$.

- Define the *rank function* of $\mathcal{V}$ by

$$r : \mathcal{P}(\mathcal{V}) \longrightarrow \mathbb{Z}\,, \qquad r(\{v_{i_1}, \ldots, v_{i_k}\}) = \dim\langle v_{i_1}, \ldots, v_{i_k}\rangle\;.$$

$|\mathcal{C}(\mathcal{H})|$ depends only on the rank function of $\mathcal{V}$.

Update graphs of SDSs

- *Question*: When does $[\mathfrak{F}_Y, \pi] = [\mathfrak{F}_Y, \sigma]$, for distinct update orders $\pi, \sigma \in S_Y$?

Definition. The *update graph* $U(Y)$ has vertex set S_Y. The edge $\{\pi, \sigma\}$ is present iff:

- π and σ different by exactly an adjacent transposition $(i,\ i+1)$,
- $\{\pi_i, \pi_{i+1}\} \notin \mathrm{e}[Y]$.

Example. Let Circ_4 be the circular graph on 4 vertices.

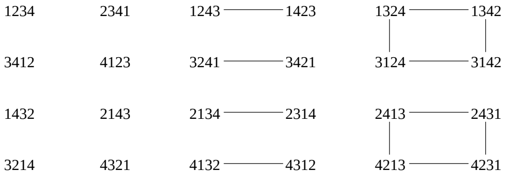


Figure: The update graph $U(\mathrm{Circ}_4)$.

Functional equivalence of SDSs

Define an equivalence relation $\sim_Y$ on S_Y by $\pi \sim_Y \sigma$ if π and σ are on the same connected component of $U(Y)$.

Prop. If $\pi \sim_Y \sigma$, then $[\mathfrak{F}_Y, \pi] = [\mathfrak{F}_Y, \sigma]$.

- An ordering $\pi \in S_Y$ induces an acyclic orientation of Y, denoted O_Y^π.
- There is a bijection between

$$f_Y \colon S_Y/{\sim_Y} \longrightarrow \mathrm{Acyc}(Y), \qquad f_Y([\pi]_Y) = O_Y^\pi \;.$$

Thus, $\alpha(Y) = |\mathrm{Acyc}(Y)|$ is an upper bound for the number of functionally distinct SDS maps $[\mathfrak{F}_Y, \pi]$ for a fixed choice of $\mathfrak{F}_Y$. This bound is known to be sharp.

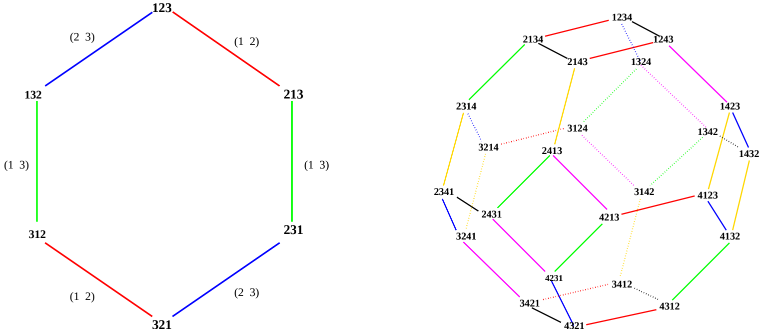
Permutahedra

The *n-permutahedron* Π_n is the convex hull of all permutations of the points $(1, 2, \ldots, n) \in \mathbb{R}^n$. It is an $(n-1)$-dimensional polytope.

The vertices and edges of Π_n can be labeled as follows:

- Two vertices are adjacent if they differ by swapping two coordinates in adjacent position.
- An edge is labeled with a transposition $(x_i,\ x_j)$ of the values of the two entries that are swapped.

Note: This labeling scheme does not agree with the geometric coordinates of the vertices!



(a) Π_3 (b) Π_4

Figure: Permutahedra, for $n = 3$ and $n = 4$.

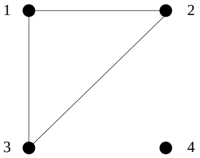
Constructing $U(Y)$ from Π_n

- Π_n is the update graph of E_n.

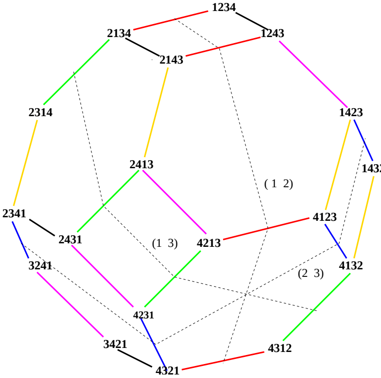
- Each transposition $(i\ j) \in S_n$ corresponds with a complete set of parallel edges of Π_n.

- The update graph $U(Y)$ can be constructed by "cutting" Π_n with a hyperplane $H^n_{i,j}$ for every edge $\{i,j\} \in \mathrm{e}[Y]$.

An example



(a) Y



(b) Constructing $U(Y)$

Figure: Hyperplanes cuts corresponding with the edges $\{1,2\}$, $\{2,3\}$, and $\{1,3\}$ in $Y < K_4$.

The rank function of a graph

Definition. The *rank function* of a graph Y is the function

$$r_Y : \mathcal{P}(\mathcal{Y}) \longrightarrow \mathbb{Z}\,, \qquad r_Y(Z) = r(\{v^n_{i,j} \mid \{i,j\} \in \mathrm{e}[Z]\})\,,$$

where $v^n_{i,j}$ is the normal vector of the hyperplane $H^n_{i,j}$ from the constructing of $U(Y)$.

Definition. Let $f_1, \ldots, f_n$ be a basis for the dual space of $\mathbb{R}^n$. For a graph $G = (V, E)$, let $\mathcal{H}(G)$ be the arrangement defined by

$$\mathcal{H}(G) = \{\ker(f_i - f_j) \mid \{i,j\} \in E\}\,.$$

$\mathcal{H}(G)$ is called the *graphic arrangement* of G.

Prop. If $Z < Y$, then $r_Y(Z)$ is the number of edges in a spanning forest of Z, i.e.,

$$r_Y(Z) = |\mathrm{v}[Z]| - n(Z)\,.$$

Cycle equivalence of SDSs

Definition Two finite dynamical systems $\phi, \psi \colon K^n \to K^n$ are *cycle equivalent* if there exists a bijection $h \colon \mathrm{Per}(\phi) \longrightarrow \mathrm{Per}(\psi)$ such that

$$\psi|_{\mathrm{Per}(\psi)} \circ h = h \circ \phi|_{\mathrm{Per}(\phi)} \ .$$

- Let $\sigma, \tau \in S_n$ be

$$\sigma = (n, n-1, \dots, 2, 1) \ , \qquad \tau = (1, n)(2, n-1) \cdots (\lceil \tfrac{n}{2} \rceil, \lfloor \tfrac{n}{2} \rfloor + 1) \ ,$$

and let C_n and D_n be the groups

$$C_n = \langle \sigma \rangle \qquad\qquad D_n = \langle \sigma, \tau \rangle \ .$$

These groups act on update orders $\pi = \pi_1 \pi_2 \cdots \pi_n$ by *shift* and *reflection* as follows:

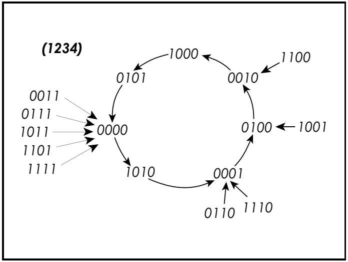
$$\boldsymbol{\sigma}(\pi) := \sigma \cdot \pi = \pi_2 \pi_3 \cdots \pi_n \pi_1 \ , \qquad \boldsymbol{\rho}(\pi) := \tau \cdot \pi = \pi_n \pi_{n-1} \cdots \pi_2 \pi_1 \ .$$

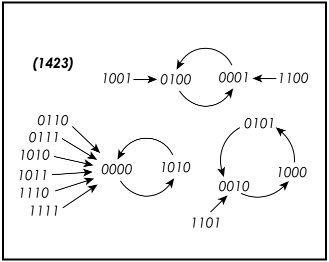
Cycle equivalence of SDSs

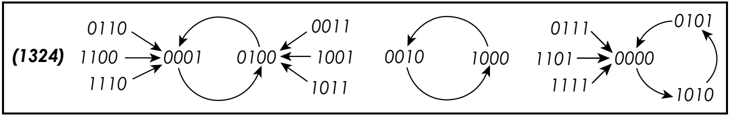
Theorem

For any $\pi \in S_Y$, the SDS *maps* $[\mathfrak{F}_Y, \pi]$ *and* $[\mathfrak{F}_Y, \boldsymbol{\sigma}(\pi)]$ *are cycle equivalent. Moreover, if* $K = \mathbb{F}_2$*, then these are cycle equivalent to* $[\mathfrak{F}_Y, \boldsymbol{\rho}(\pi)]$ *as well.*

Example. $[\mathbf{Nor}_{\mathrm{Circ}_4}, \pi]$ for given update sequences:







Enumeration of equivalence classes of SDS maps

- $\alpha(Y) := \text{Acyc}(Y) = T_Y(2,0)$ satisfies

$$\alpha(Y) = \alpha(Y/e) + \alpha(Y \setminus e) \qquad \text{for any edge } e$$

  and is an upper bound for the number of SDS maps $[\mathfrak{F}_Y, \pi]$ for a fixed choice of $\mathfrak{F}_Y$ *up to functional equivalence*.

- $\kappa(Y) := \text{Acyc}(Y) = T_Y(1,0)$ satisfies

$$\kappa(Y) = \kappa(Y/e) + \kappa(Y \setminus e) \qquad \text{for any cycle edge } e$$

  and is an upper bound for the number of SDS maps $[\mathfrak{F}_Y, \pi]$ for a fixed choice of $\mathfrak{F}_Y$ *up to cycle equivalence* [7].

Remark: $\kappa(Y)$ is also a sharp upper bound for the number of conjugacy classes of Coxeter elements of a simply-laced Coxeter system with Coxeter graph Y (see [8, 9]).

Reachable attractor basins as a measure of stability

Consider a sequential dynamical system over a graph Y, with 2-threshold functions.

▶ If Y is a tree, then there are states that can reach $2^n - n = O(2^n)$ fixed points by varying the update order.

▶ If $Y = K_n$, then any given state can reach at most $n + 1$ fixed points by varying the update order.

▶ This can be extended to $G_{n,p}$ for various values of p (see [4]).

Conclusion. In general, adding edges to the dependency graph of an SDS causes the dynamics to become *more* stable with respect to update order.

Cycle equivalence as a measure of stability

Prop. If Y is a tree, then $\kappa(Y) = 1$.

Corollary. If Y is a tree, then for a fixed choice of functions $\mathfrak{F}_Y$, all SDS maps $[\mathfrak{F}_Y, \pi]$ are cycle equivalent.

The functions $\alpha(Y)$ and $\kappa(Y)$ are a measure of system complexity, and are Tutte-Grothendieck invariants.

▶ Adding more edges to Y can only increase $\alpha(Y)$ and $\kappa(Y)$, and thus the number of possible SDS maps, and cycle structures of the maps.

Conclusion. In general, adding edges to the dependency graph of an SDS causes the dynamics to become *less* stable with respect to update order.

What is going on???

We are discussing *different notions* of update order instability!

By update order stability, one can measure:

- How many different possible cycles structures (long term behaviors) are there. . .
- How many different attractor basins can be reached from a particular state . . .
- Or something else! (e.g., which states can arise as fixed points [5, 6]) . . .

. . . as the update order is perturbed.

Moral: Be *careful* when making general statements about the update order stability of a dynamical system!

Question: These ideas have only been studied independently. Is there a way to tie them together to paint a clearer picture?

Application – Role of the graph structure in complex systems

When studying complex systems, it is natural to ask the following questions:

Question 1. *What role does the structure of the dependency graph play in the dynamics of the a system defined over it?*

Question 2. *Does there exist a good measure, or classification of graphs, useful for people studying graph dynamical systems?*

Key idea: Such a measure should be able to capture the cycle structure of the graph.

Measures such as the *degree distribution* or *clustering coefficient* can't detect the presence of large cycles.

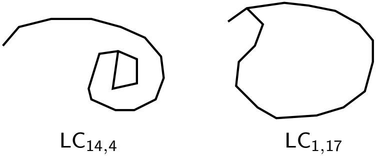


Figure: Two graphs with the same degree distribution and clustering coefficient, but "dynamically different."

Approach: Edge shattering

▶ Recall that the rank function r_Y captures the cycle structure of a graph, and is determined by how many components remain upon removal of a certain subset of edges.

▶ For a graph with m edges, the rank function is of size $\Theta(2^m)$, thus it is uncomputable for most graphs.

▶ However, we can extract useful *edge shattering properties* from the following functions $[0,1] \to \mathbb{N}$:

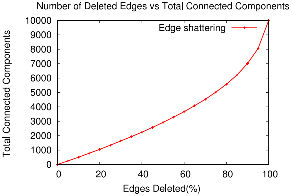
- $\mu_Y(k)$: average number of components when k% of edges are removed from Y.
- $M_Y(k)$: maximum number of components when k% of edges are removed from Y.
- $\lambda_Y(k)$: average size of largest component when k% of edges are removed from Y.
- $\sigma_Y(k)$: average size of a component when k% of edges are removed from Y.

Questions and Future Research

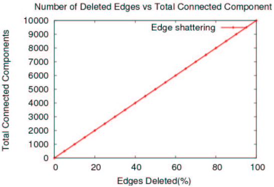
▶ How well do these edge shattering functions distinguish commonly studied classes of graphs? For example:

- Classical random graphs, such as $G_{n,p}$ and $G_{n,M}$.
- Small-world networks (Watts & Strogatz, 1998)
- Scale-free networks (Barabási and Albert, 1999)
- Real-world biological, epidemiological, and social networks.

▶ Can one re-construct a network of a particular size that satisfies certain edge shattering properties (hard!).



(a) $G_{n,p}$ for $n = 10^4$, $p = 10^{-3}$



(b) A 10000-vertex tree

Figure: Edge shattering plots of σ_Y for two graphs

References

F. D. J. Dunstan. **Matroids and Submodular Functions**. *The Quarterly Journal of Mathematics*. 27 (3) 1976:339–348.

U. Karaoz, T. M. Murali, S. Letovsky, Y. Zheng, C. Ding, C. R. Cantor, S. Kasif. **Whole-genome annotation by using evidence integration in functional-linkage networks**. *Proc. Natl. Acad. Sci.*, 101:2888–2893, 2004.

R. Laubenbacher, A. S. Jarrah, H. S. Mortveit, S. S. Ravi. **A mathematical formalism for agent-based modeling**. *Encyclopedia of Complexity and System Science*. Springer Verlag, 2008.

V. S. A. Kumar, M. Macauley, H. S. Mortveit. **Update order instability in graph dynamical systems**. In progress.

M. Macauley, J. McCammond, and H. S. Mortveit. **Order independence in asynchronous cellular automata**. *J. Cell. Autom*. 3 (1), 2008:37–56. arXiv:0707.2360.

M. Macauley, J. McCammond, and H. S. Mortveit. **Dynamics groups of finite asynchronous cellular automata**. Submitted to *Found. Computat. Math.*, 2008. arXiv:0808.1238.

M. Macauley, H. S. Mortveit. **Cycle equivalence of graph dynamical systems**. *Nonlinearity*. To appear, 2009. arXiv:0802.4412.

M. Macauley, H. S. Mortveit. **On enumeration of conjugacy classes of Coxeter elements**. *Proc. AMS*. 136, 2008:4157–4165. arXiv:0711.1140.

M. Macauley, H. S. Mortveit. **A solution to the conjugacy problem for Coxeter elements in simply laced Coxeter groups**. Submitted to *Adv. Math.*, 2008.

H. S. Mortveit, C. M. Reidys. **An introduction to sequential dynamical systems**. Springer Verlag, 2007.

P. Orlik, H. Terao. **Arrangements of hyperplanes**. A Series of Comprehensive Studies in Mathematics. Spring Verlag, 1992.

G. M. Ziegler. **Lecture on Polytopes**. Vol. 152 of Graduate Texts in Mathematics, Springer Verlag, 1995.