Chemical Reaction Networks

Part 1: Chemical Reaction Networks — what, where, why?

- What are Chemical Reaction Networks?
- Where can you use them?
- Why they are hard to analyze?
- Why they are interesting, mathematically and practically?
- Levels at which one can use them as a modeling framework

What are Chemical Reaction Networks (loose)

- Models of chemical systems + much more
- Discrete-state Markovian stochastic processes
- Topologically interesting graph systems
- Laboratories in which to study:
  - the consequences of *concurrency* in transformations
  - how topology controls dynamics or possibilities for order

What are Chemical Reaction Networks (tighter)

- Sets of reactions that act on sets of species and:
  - reactions can fire independently and randomly
  - convert sets of reactants to sets of products all at once

- Hopping on simple graphs is a degenerate case

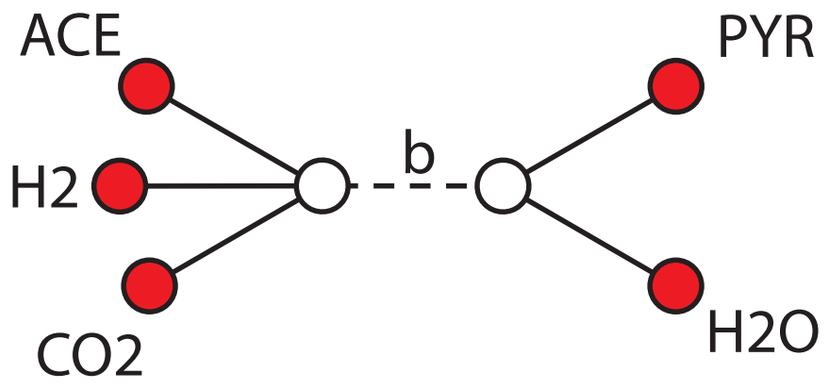
- More generally one requires multi-hypergraphs
  (we will represent these with doubly-bipartite graphs)

Stochastic processes with or without concurrency

- Stochastic process on a simple graph:

  - nodes are “all alike”; particles can occupy anywhere
  - particles hop over links — simple diffusion on the graph

- CRN: reactions are conditional on all inputs’ being available and all outputs’ being formed



- Reactions are independent like those on the simple graph
  But moves in the state space are not independent

Pause to emphasize what is the MAIN POINT

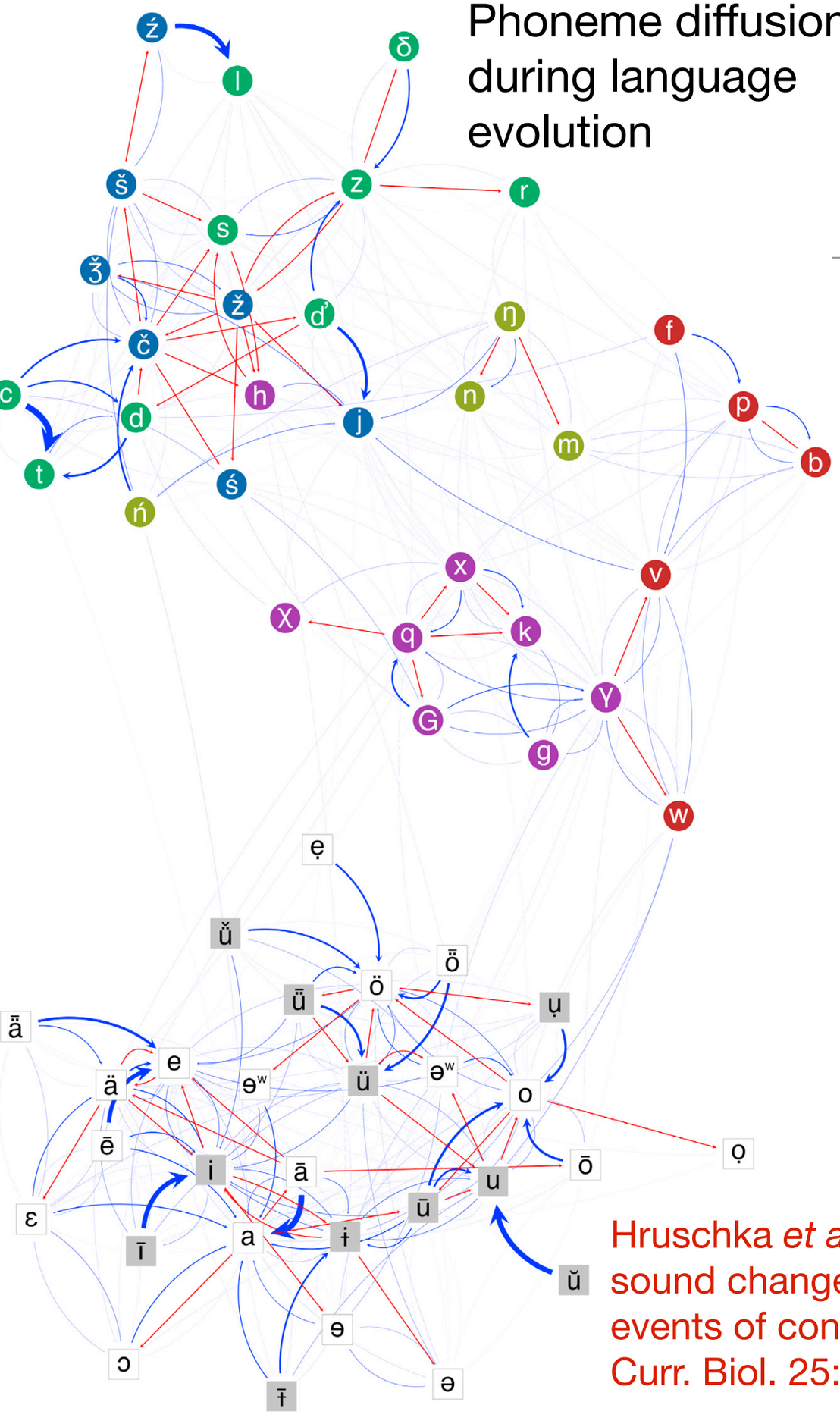
- It is the ***species*** that carry state information from one moment to the next in a Markov CRN

- For simple graphs, events that change state are independent species-by-species

- For non-simple hypergraphs, independence is lifted to the reaction level, but may not hold species-by-species

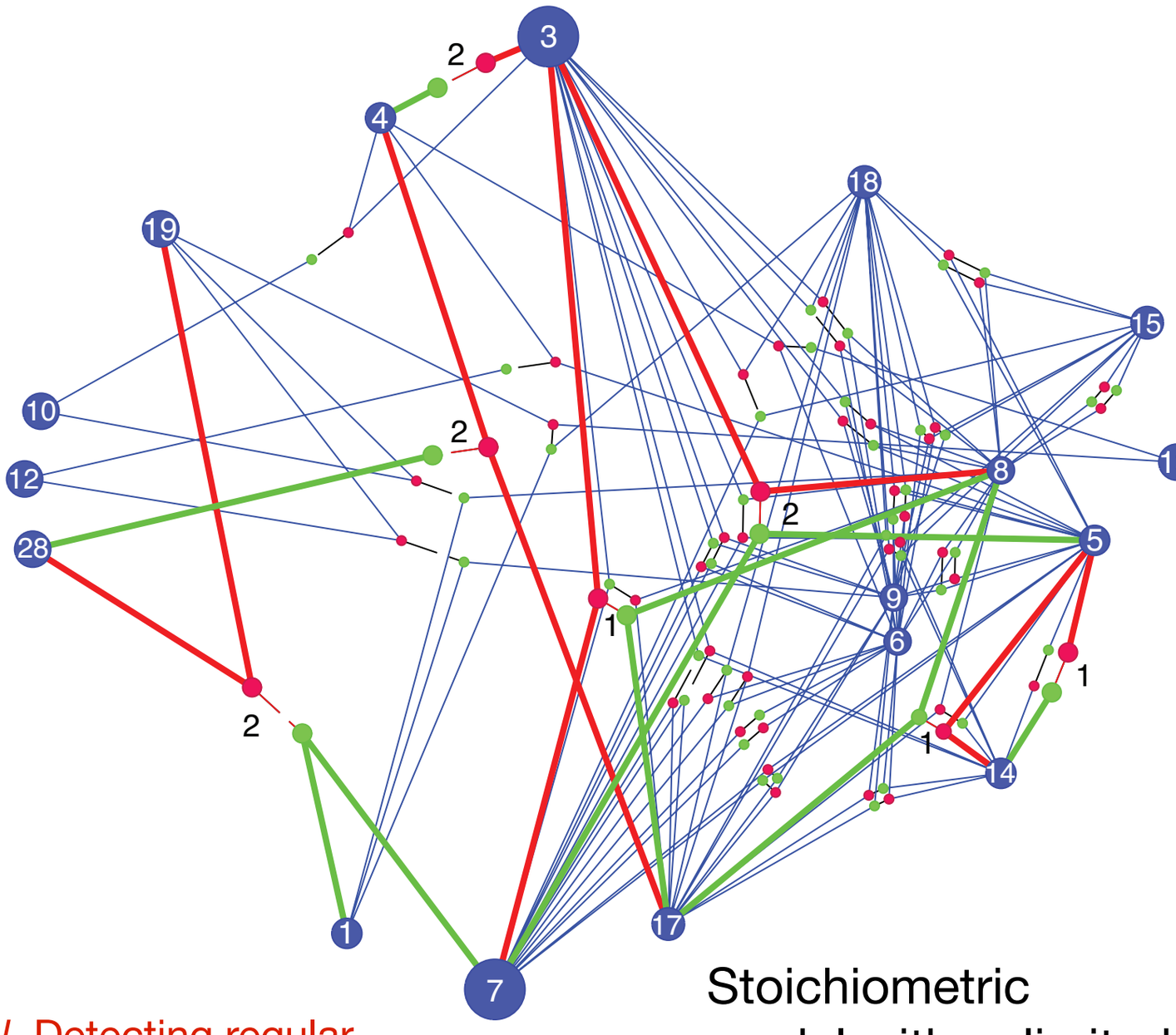
- This difference is the source of complexity

What might one model with CRNs?



Phoneme diffusion during language evolution

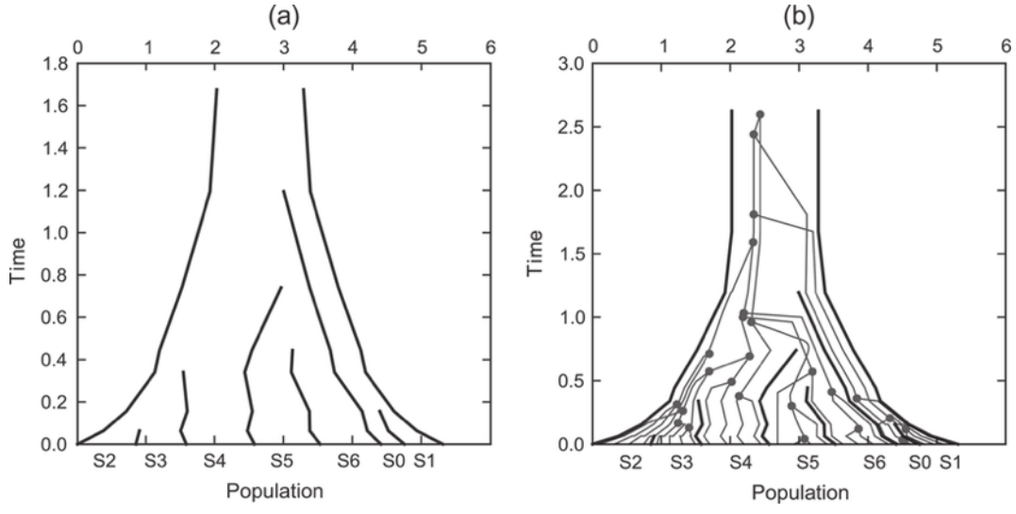
Hruschka *et al.* Detecting regular sound changes in linguistics as events of concerted evolution Curr. Biol. 25:1--9 (2015)



Stoichiometric model with unlimited solutions to a fixed conversion problem

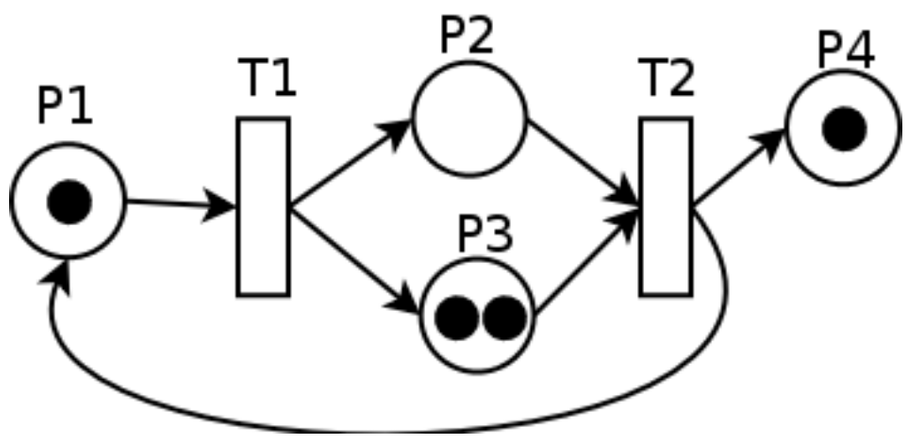
Other applications and other names for CRNs and closely-related systems

- Population processes such as birth-death processes, including evolutionary games and lots of genetics cases



www.researchgate.net/publication/38083310

- In computer science the same concept of concurrency is captured by Petri Nets for process calculus



Levels at which CRN models can be defined and studied

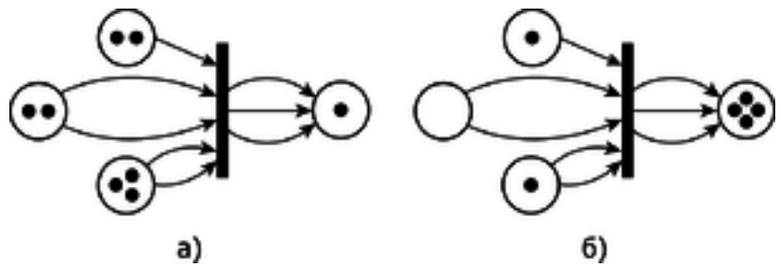
- Chemical rate equation

$$\frac{dn}{dt} = Y\,\mathbb{A}\,\psi_Y$$

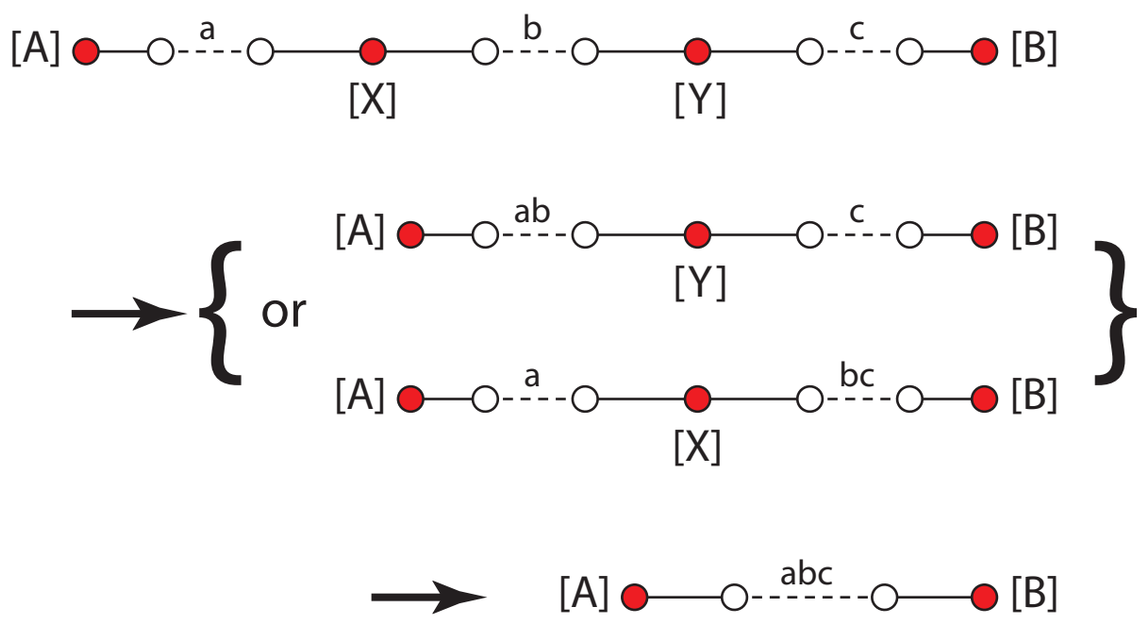
- Stochastic process for counts

$$\frac{d}{dt}\rho_{\mathrm{n}} = \sum_{\mathrm{n}'} \mathrm{T}_{\mathrm{nn}'}\rho_{\mathrm{n}'}$$

- "Petri-net proof" counting individuals

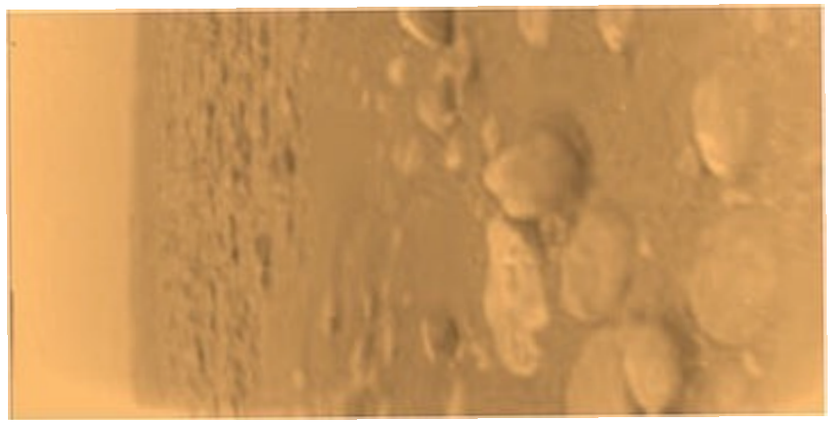


- Category-theoretic CRN-composition (J. Baez & students)

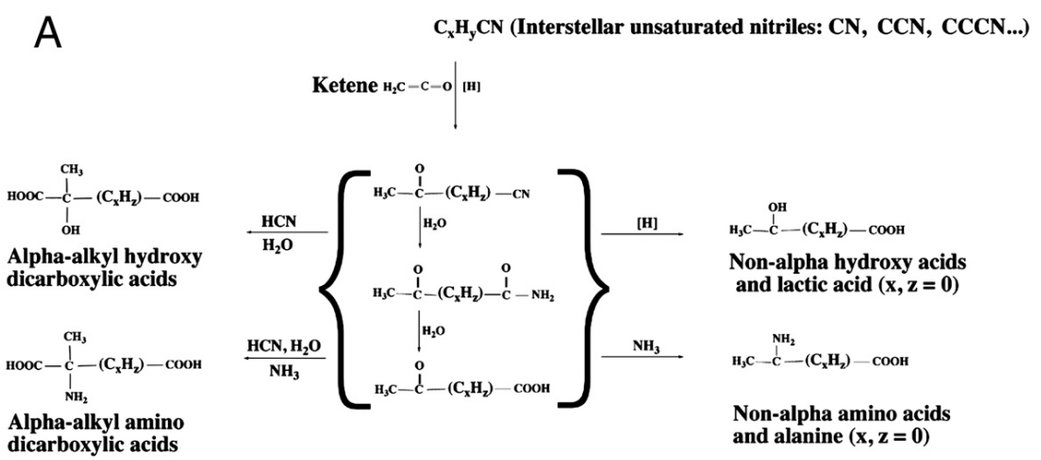


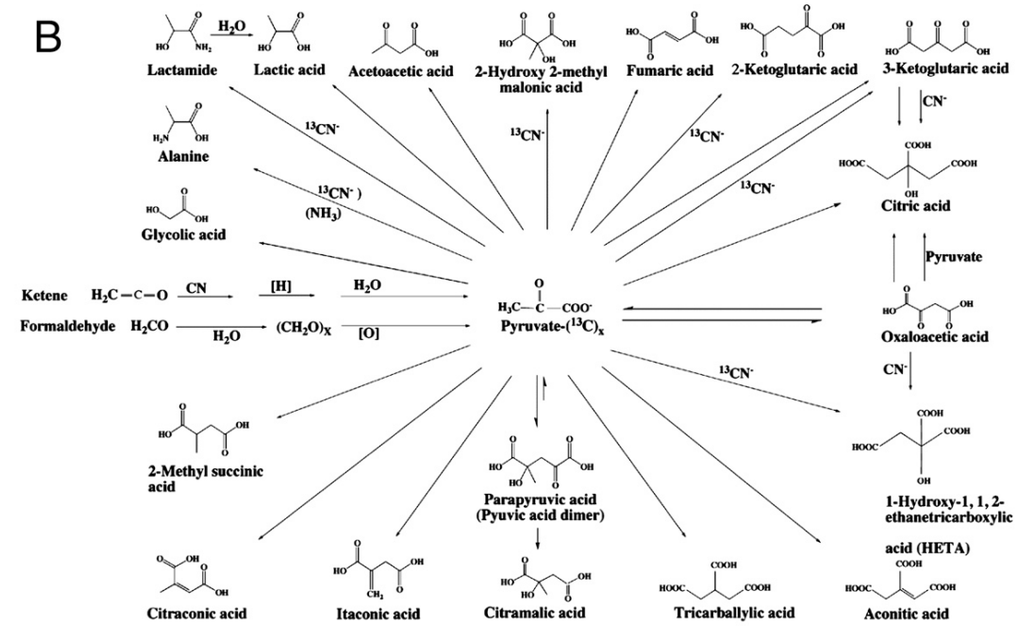
Graph-grammar methods to generate CRNs, and associated analysis tools for them

- Why do this?
- Actual networks can be too large to write down (Murchison, Titan atmosphere), or too large to even know (OOL as an example)



Huygens image: credit NASA



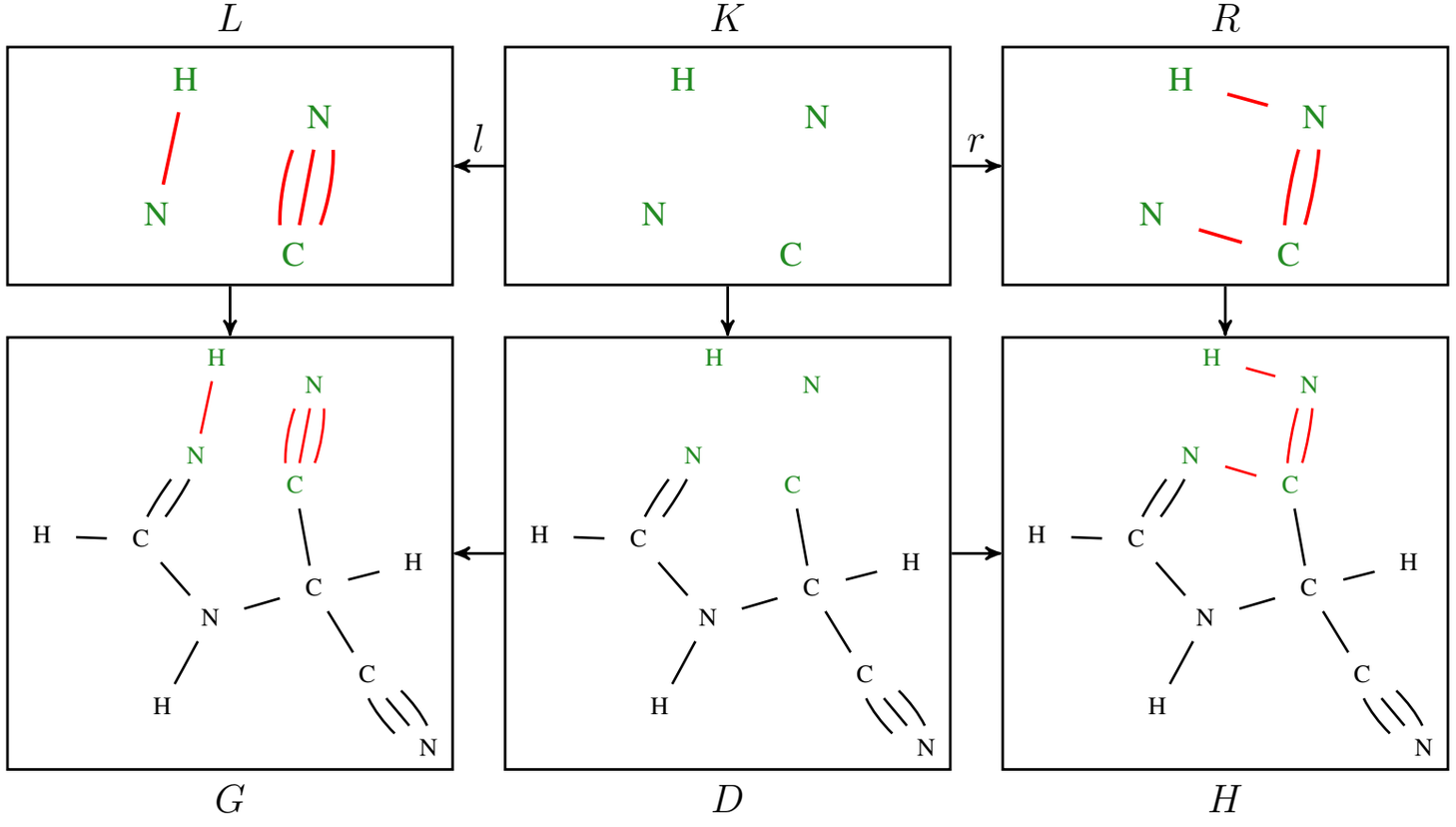


Cooper et al. PNAS 108:14051 (2011)

- Graph-grammars allow you to express an indefinitely-large network recursively from a finite set of generators

Example of a graph, graph fragment, common pattern, and rewrite rule

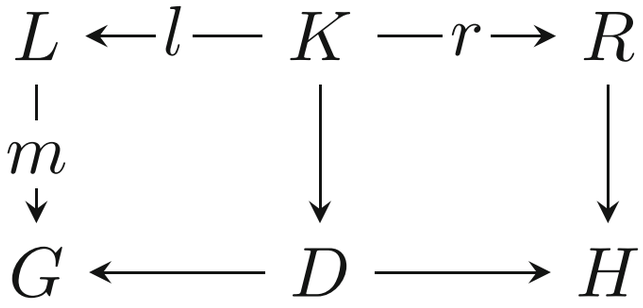
- Molecules themselves are written as graphs
- Maps among sets of molecules become graph morphisms in category theory
- Reaction mechanisms are graph-fragment morphisms
- Reaction application is a graph-embedding map



- Key point: finite generator set can produce unlimited CRNs

The double-pushout formalism for graph re-write rules

- The important thing to say here properly is that there is a commutativity requirement which defines these morphisms

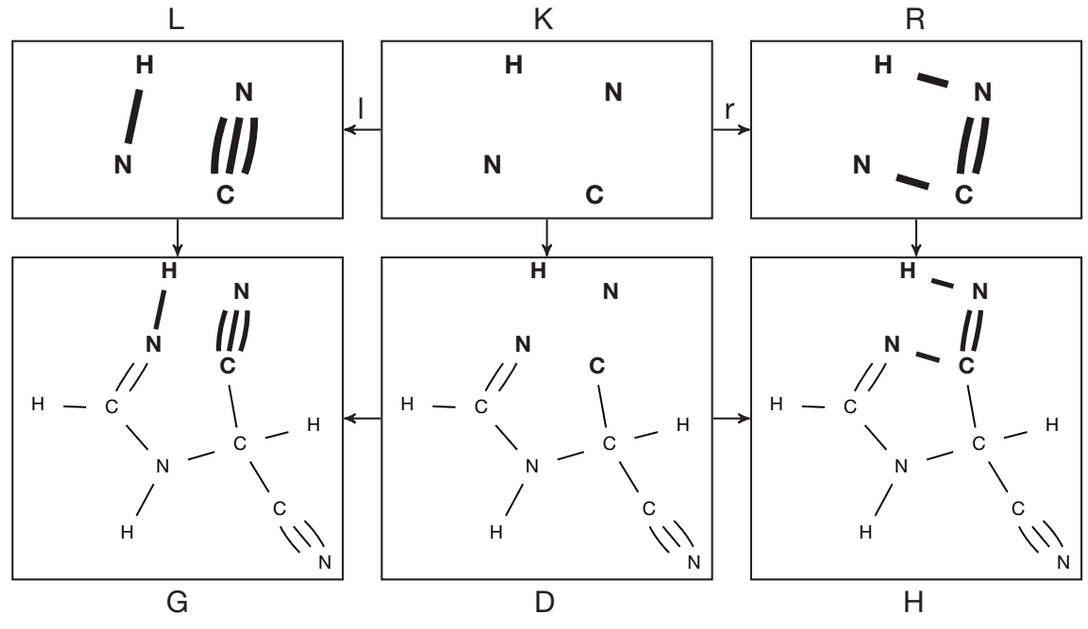


l — left pushout
(adds bonds to form L)

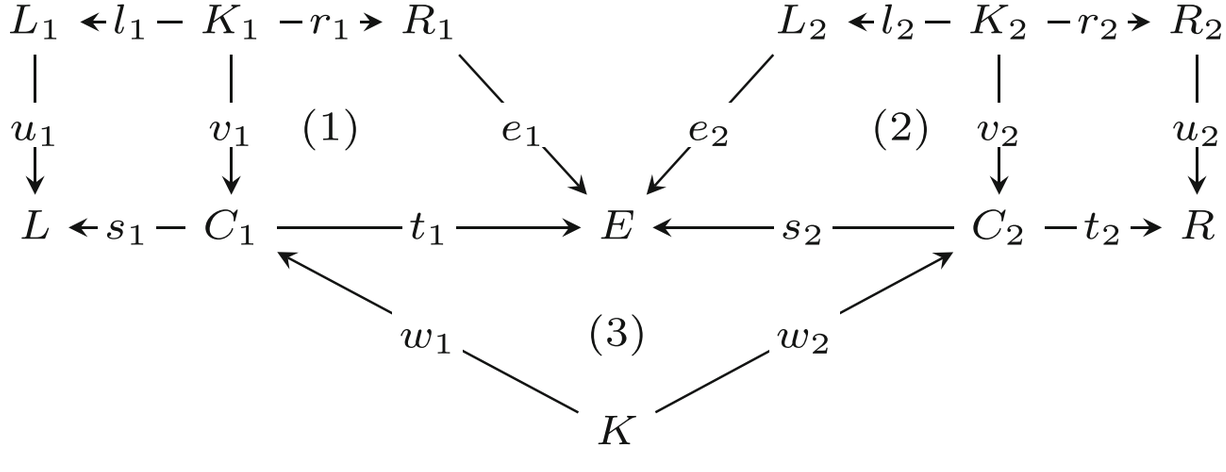
r — right pushout
(adds bonds to form R)

m — matching morphism

D — gluing condition



- Morphisms admit partial or full rule composition — think of β-reduction in λ-calculus, or Curried functions in Haskell

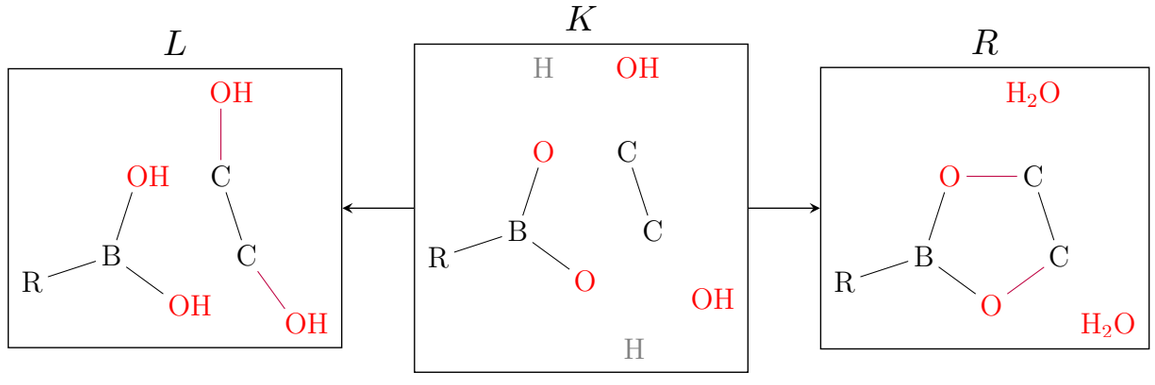


Dependency graph for rule composition

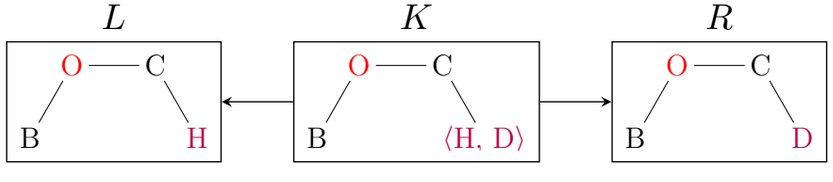
Some examples of the kinds of applied problems graph-grammars have been used to solve

- Steven Benner's pruning of the formose network with Borate to yield Ribose

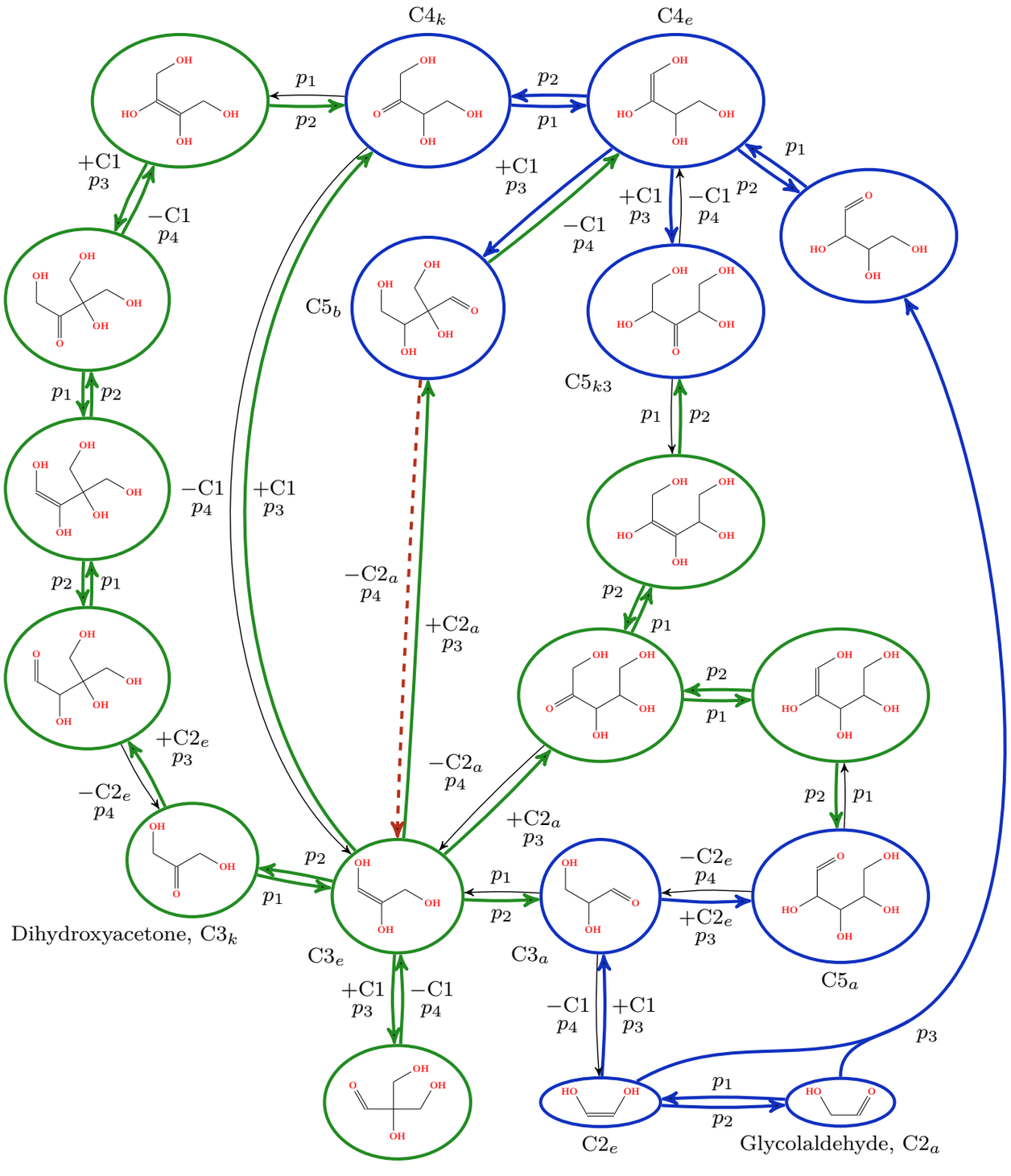
(a) Borate



(b) Borate + 1,2-diol reaction pattern, "addBorate"

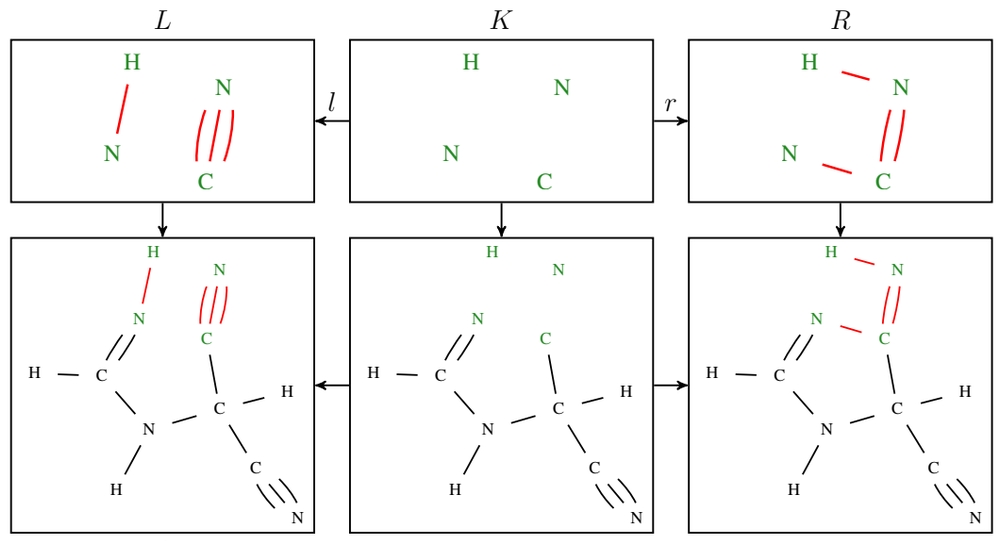


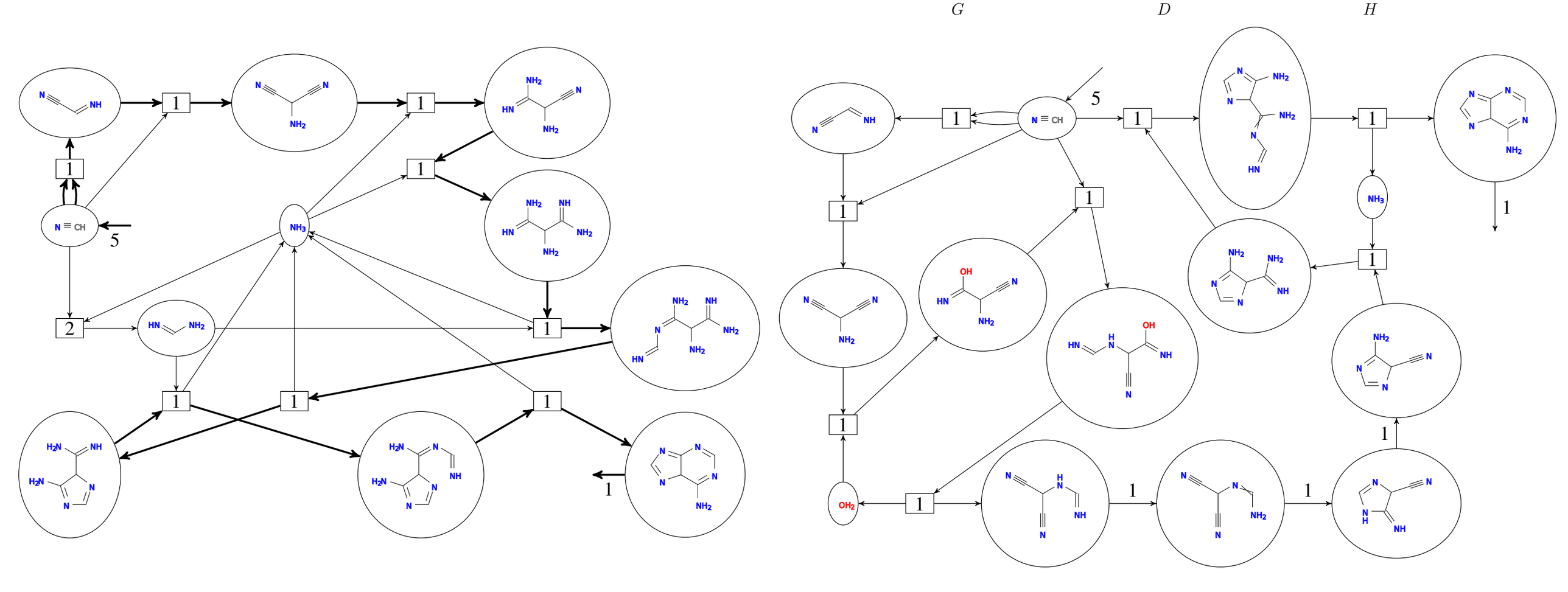
(c) Relabelling of hydrogen to make it non-reactive, "HtoD"



Some examples of the kinds of applied problems graph-grammars have been used to solve

- A network for HCN polymerization calibrated against Mass Spectroscopy molecule concentrations





Another example application

- Albert Eschenmoser's glyoxylate scenario relating HCN polymerization and hydrolysis to TCA-cycle metabolites

Original sketch by Eschenmoser

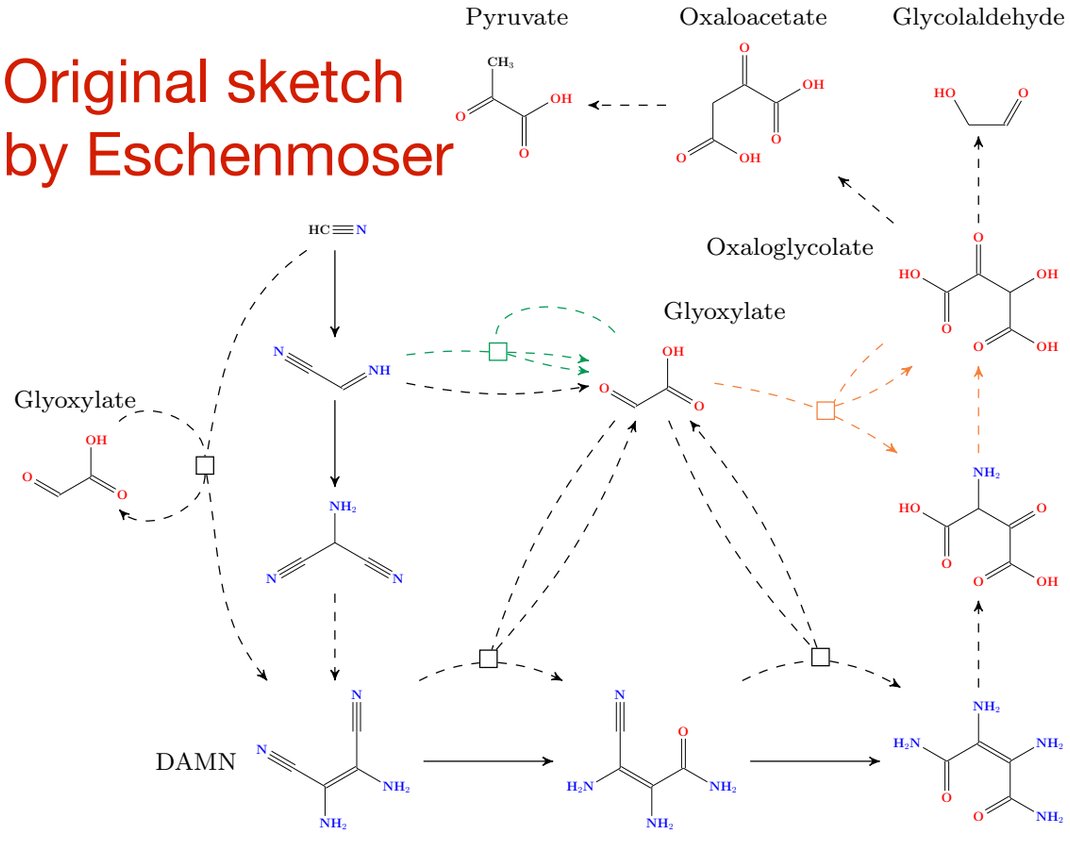
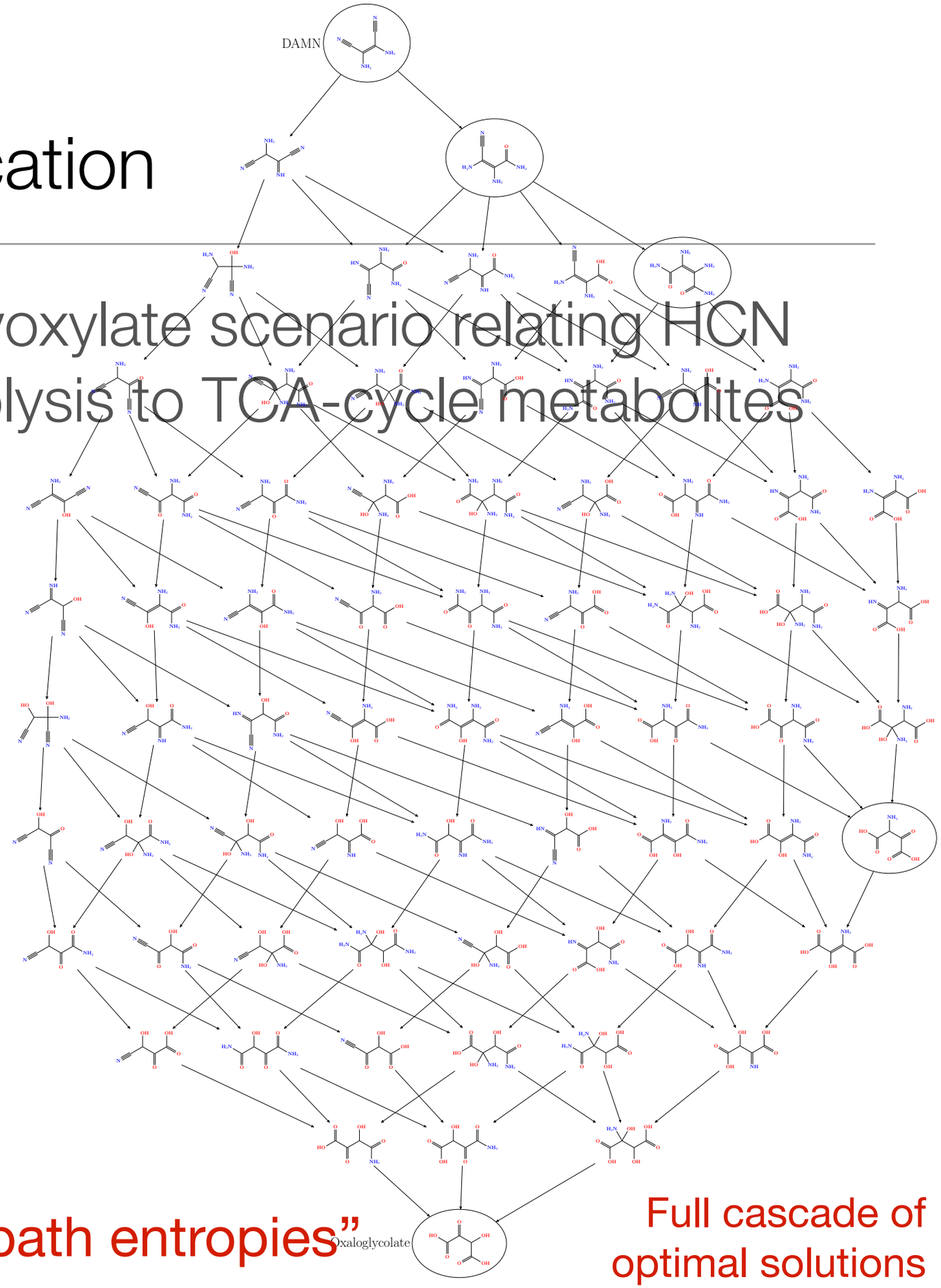


Figure 2: Overview of Eschenmoser's hypothetical chemical space.



This work suggests a role for "path entropies"

Full cascade of optimal solutions

The MedØlDatschgerl (MØD) package including a Live Playground

http://cheminf.imada.sdu.dk/mod/

- Software package for graph-based cheminformatics

- Even more important: a well-defined representation system for what constitutes a molecule / reaction / pathway

- Tools for hard network search / optimization problems

SDU University of Southern Denmark IMADA

MedØlDatschgerl

Overview

MedØlDatschgerl (MØD) is a software package developed for graph-based cheminformatics. It includes a general graph transformation system for automatically generating reaction networks from graph grammar formulations of chemistries.

The software is primarily implemented in C++, but the package includes comprehensive Python 3 binding that provides easy access most functionality. The package also includes a large visualisation module that makes it possible to automatically visualise molecules, reactions, and reaction networks. Examples of how to use the Python interface and the visualisation capabilities can be seen in the examples section, and they can all be accessed interactively in the Live Playground.

Source Code and Documentation

Each release is available at GitHub. Please also use GitHub for reporting bugs, suggesting features, and contributing code. The documentation can be found at the GitHub Page.

References

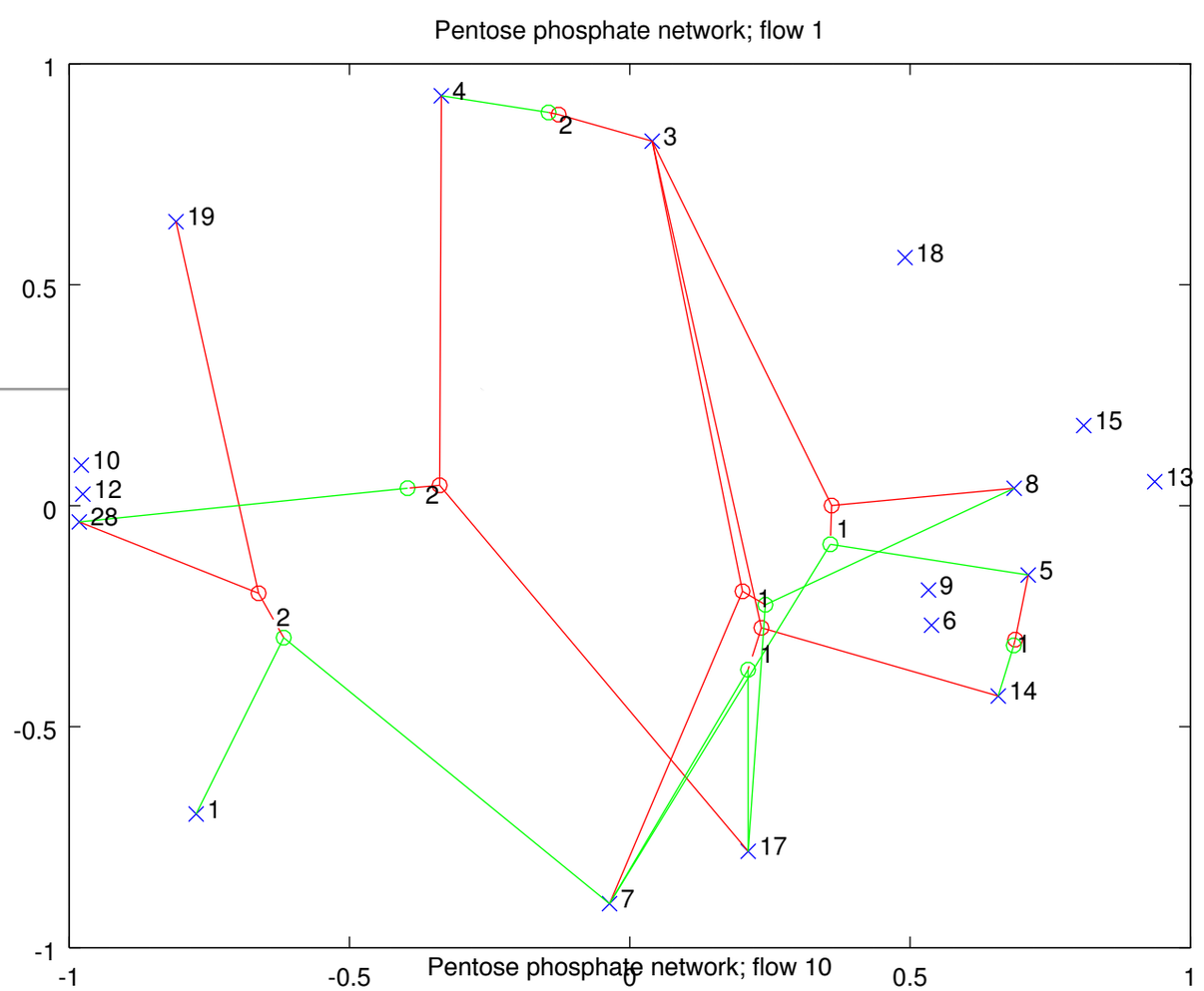
If you use MØD in your research, you may want to cite some of the following papers. You may also be interested in the Graph Grammar Library, which has been used in early versions of MØD.

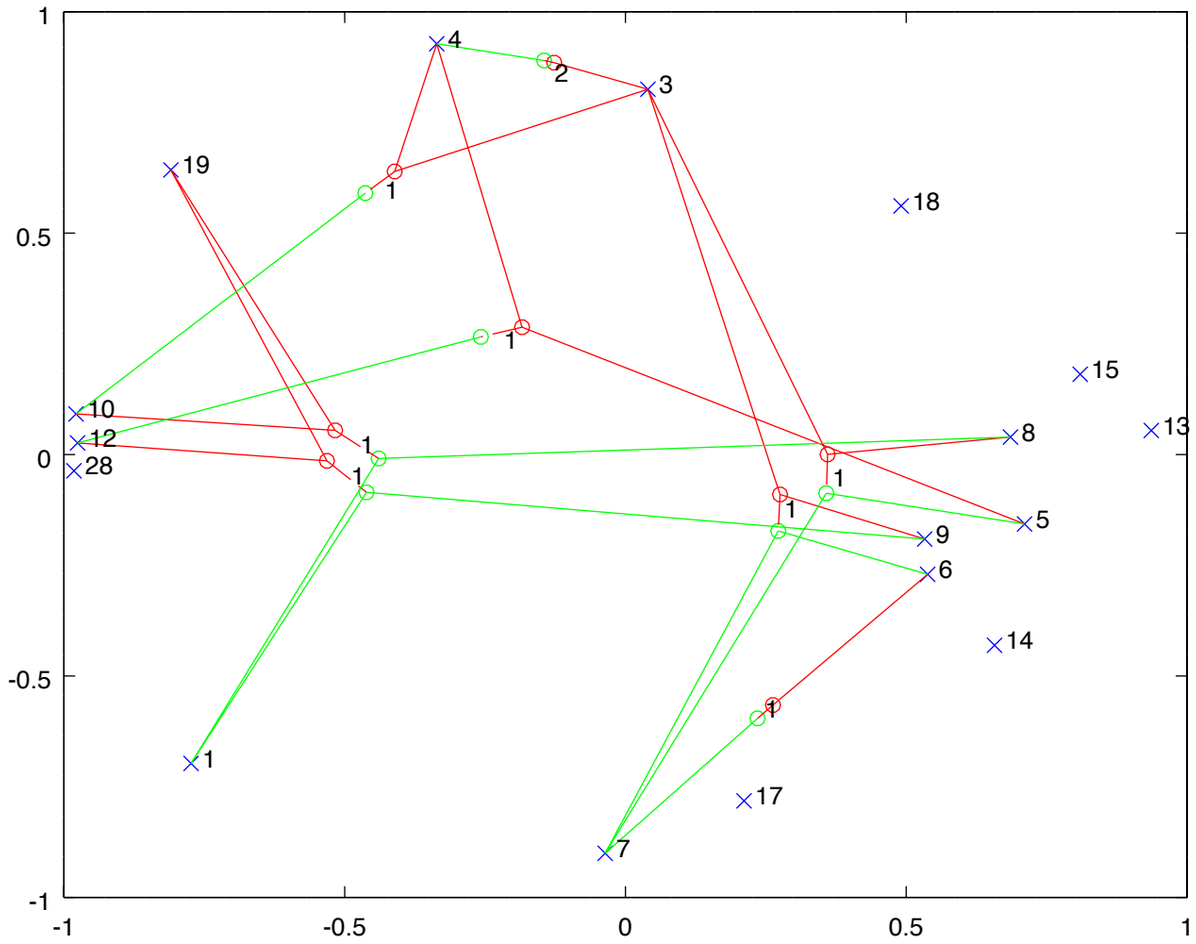
(As `BibTeX`)

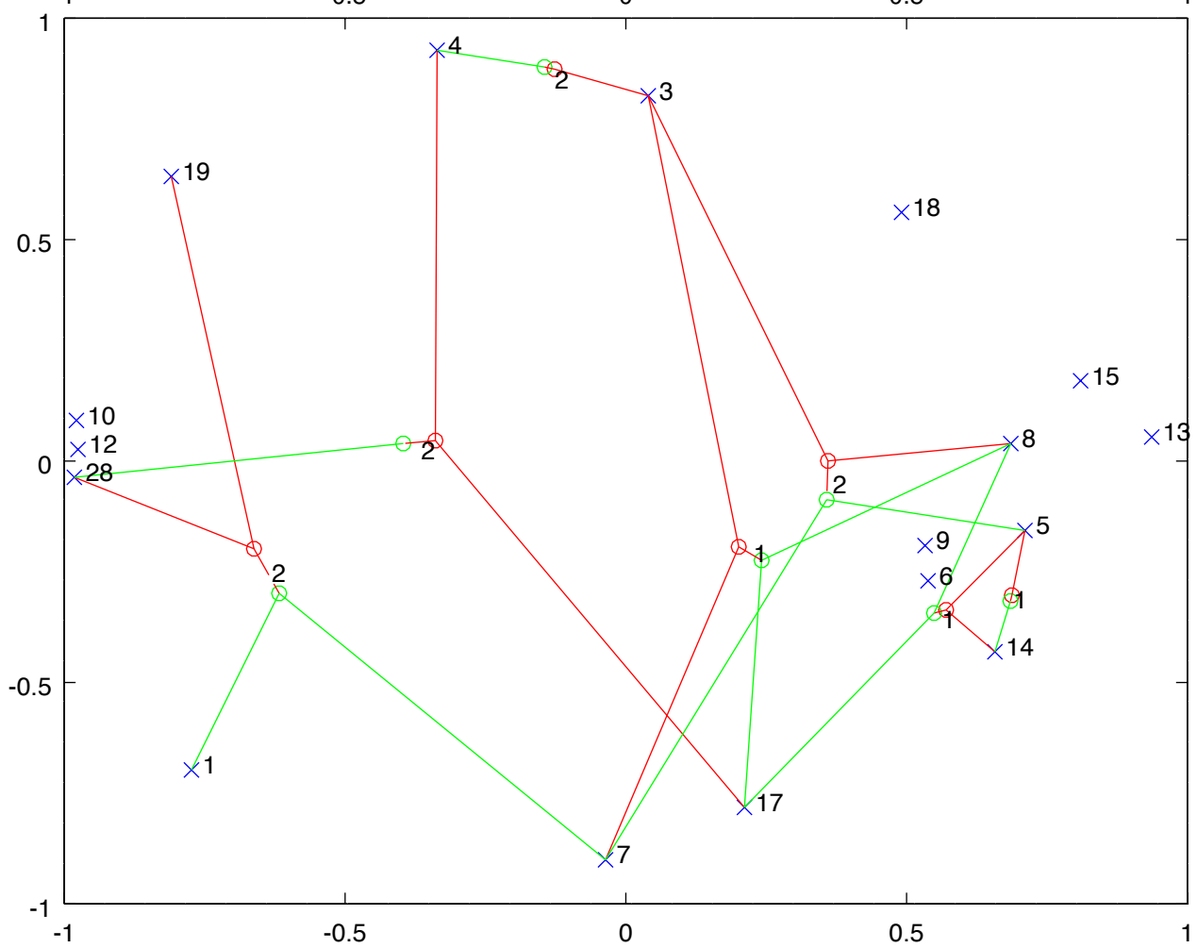
- **A Software Package for Chemically Inspired Graph Transformation**
  Jakob L. Andersen, Christoph Flamm, Daniel Merkle, and Peter F. Stadler. *Graph Transformation - 9th International Conference, ICGT 2016*, 73-88, 2016 [DOI | TR]
- **In silico Support for Eschenmoser's Glyoxylate Scenario**
  Jakob L. Andersen, Christoph Flamm, Daniel Merkle, and Peter F. Stadler. *Israel Journal of Chemistry*, 55(8):919-933, 2015. [DOI | TR]
- **50 Shades of Rule Composition — From Chemical Reactions to Higher Levels of Abstraction**
  Jakob L. Andersen, Christoph Flamm, Daniel Merkle, and Peter F. Stadler. *Formal Methods in Macro-Biology*, 8738:117-135, 2014. [DOI]
- Conference version: **Towards an Optimal DNA-Templated Molecular Assembler**
  Jakob L. Andersen, Christoph Flamm, Martin M. Hanczyc, and Daniel Merkle. *ALIFE 14: The Fourteenth Conference on the Synthesis and Simulation of Living Systems*, 14:557-564, 2014. [DOI | http]
  Journal version: **Towards Optimal DNA-Templated Computing**
  Jakob L. Andersen, Christoph Flamm, Martin M. Hanczyc, and Daniel Merkle. *International Journal of Unconventional Computing*, 11(3-4):185-203, 2015. [http]
- **Generic Strategies for Chemical Space Exploration**
  Jakob L. Andersen, Christoph Flamm, Daniel Merkle, and Peter F. Stadler. *International Journal of Computational Biology and Drug Design*, 7(2/3):225-258, 2014. [DOI | TR]
- **Navigating the Chemical Space of HCN Polymerization and Hydrolysis: Guiding Graph Grammars by Mass Spectrometry Data.**
  Jakob L. Andersen, Tommy Andersen, Christoph Flamm, Martin M. Hanczyc, Daniel Merkle, and Peter F. Stadler. *Entropy*, 15(10):4066-4083, 2013. [DOI | http]
- **Inferring chemical reaction patterns using rule composition in graph grammars.**
  Jakob L. Andersen, Christoph Flamm, Daniel Merkle, and Peter F. Stadler. *Journal of Systems Chemistry*, 4(1):4, 2013. [DOI | http]
- **Maximizing output and recognizing autocatalysis in chemical reaction networks is NP-complete.**
  Jakob L. Andersen, Christoph Flamm, Daniel Merkle, and Peter F. Stadler. *Journal of Systems Chemistry*, 3(1):1, 2012. [DOI | http]

The concept of Balanced Integer Hyperflows

- What combined flux rates balance concentrations at all species? Integer Linear Programming problem

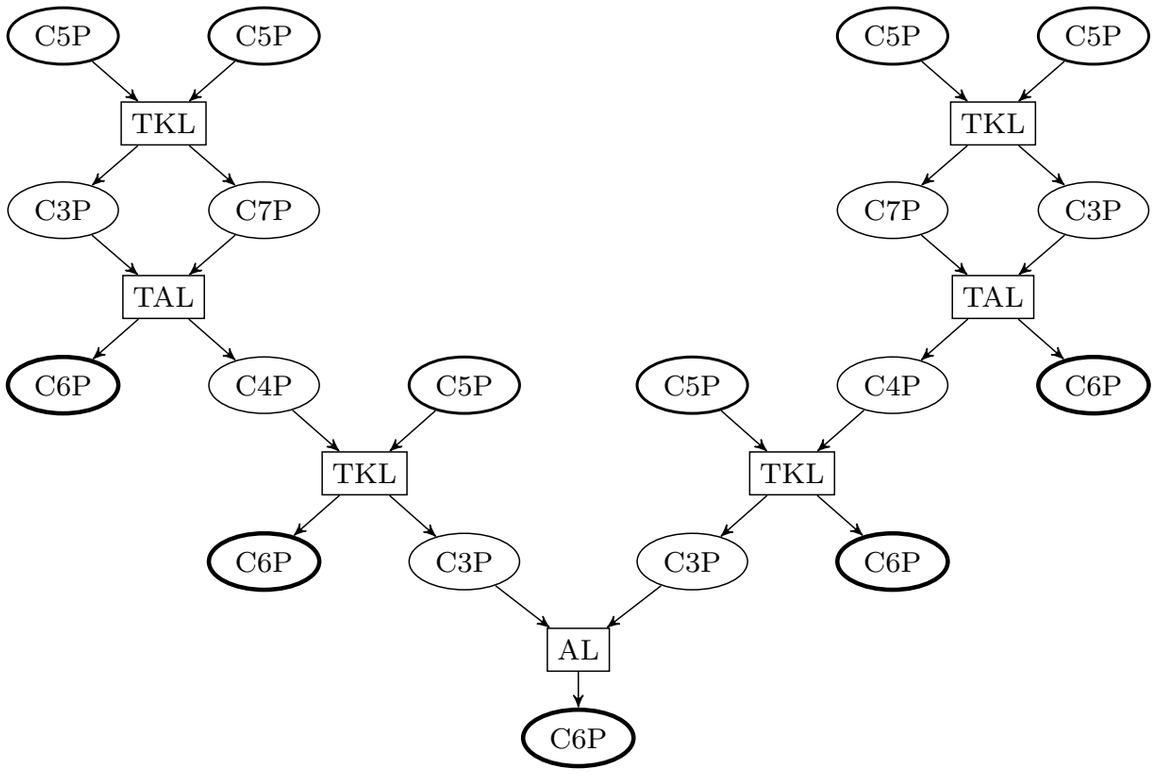




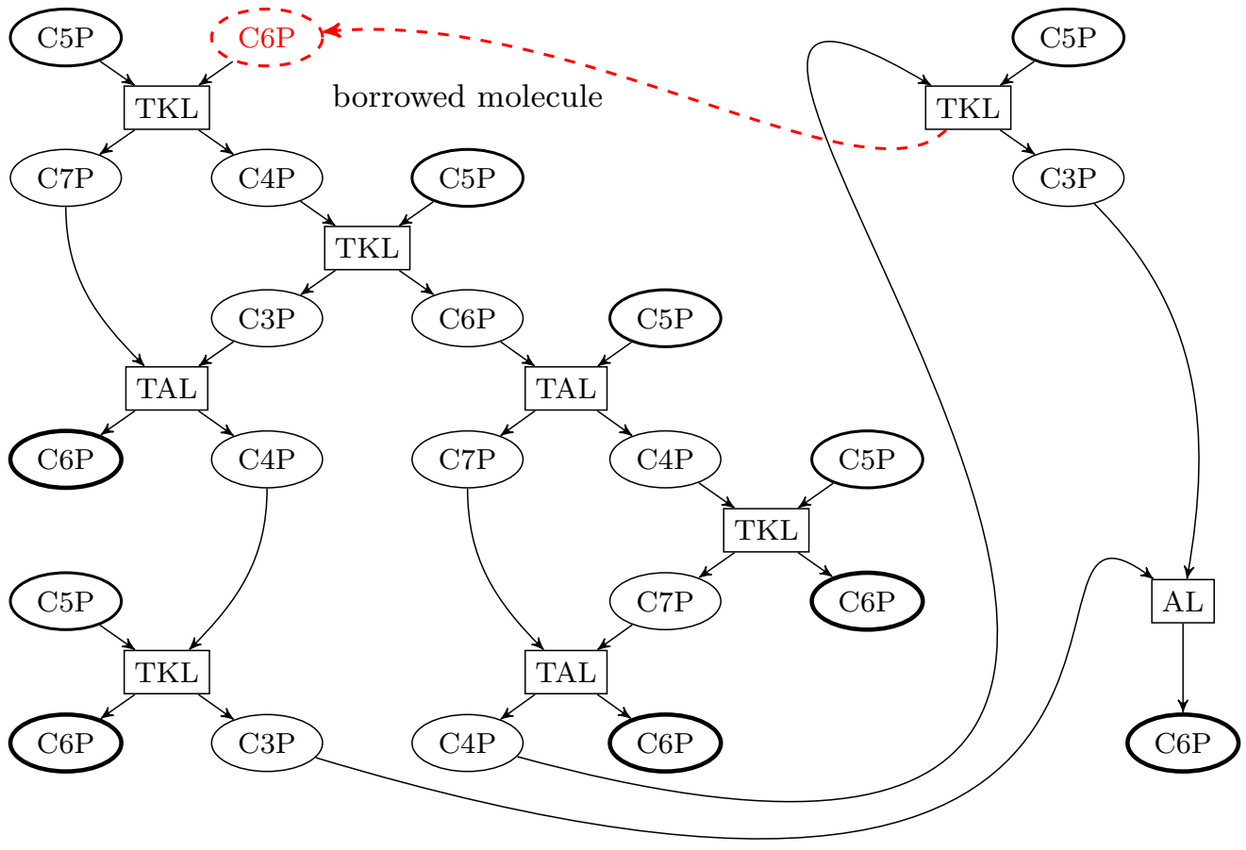


One finer level than hyperflows: Interpretations (also known as “Petri net proofs”)

Net conversion in the integer hyperflow: 6C5 -> 5C6



A realizable pathway



An un-realizable pathway
unless a molecule is borrowed

The hard math that a good computational platform helps you do

- Equivalence of molecules under different pathways: must solve graph isomorphism problem very many times

- Many search and optimization problems that are low-order polynomial on simple graphs are NP complete on hypergraphs Andersen *et al.* J. Sys. Chem. 3:1 (2012)

- For brute-force networks, need Integer-Linear-Programming help

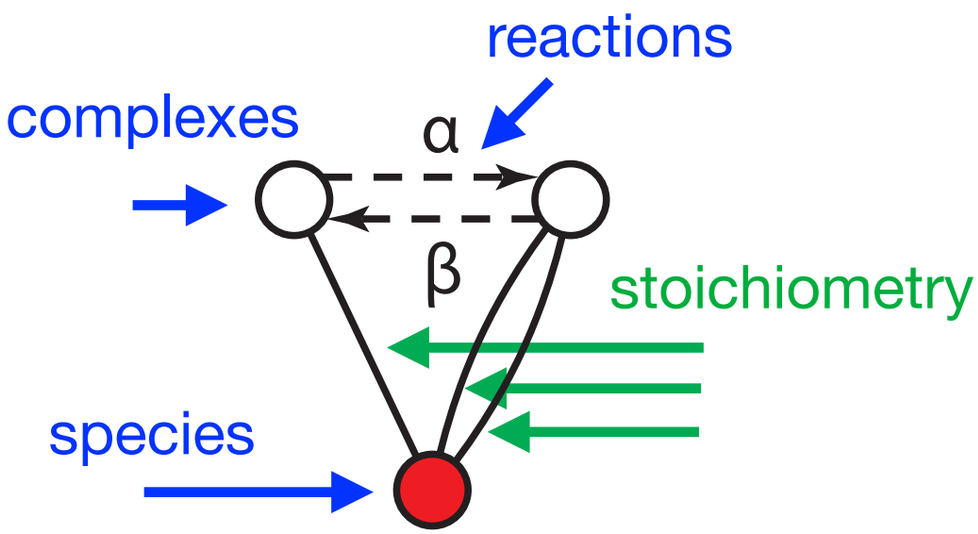
- For networks beyond brute force, good heuristics become essential

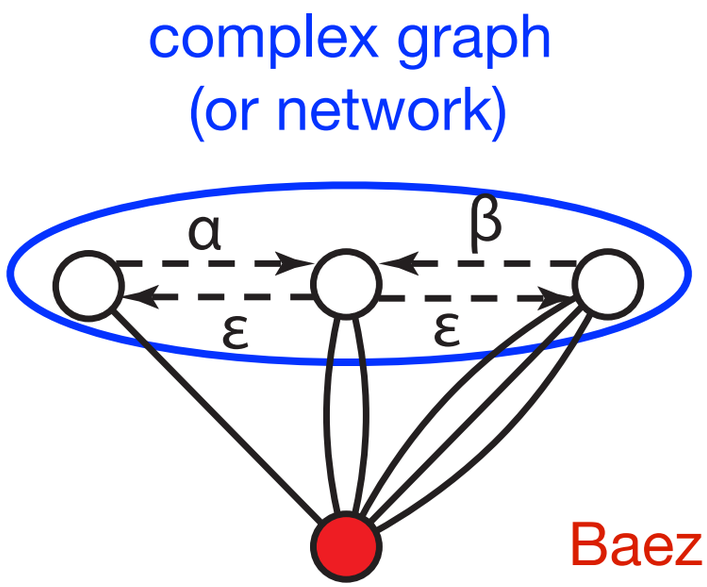
Part 2: The discrete math (topology, graph theory) of CRNs

- Main concepts, terminology, and graphical notations for CRNs

- The largest component of a CRN on which dynamics behaves the same way as a random walk on a simple graph

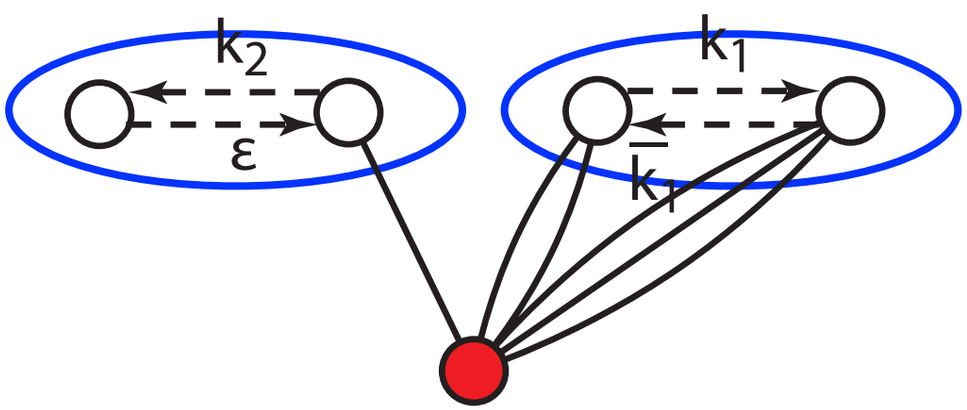
CRN basics: terms and concepts (graphics)

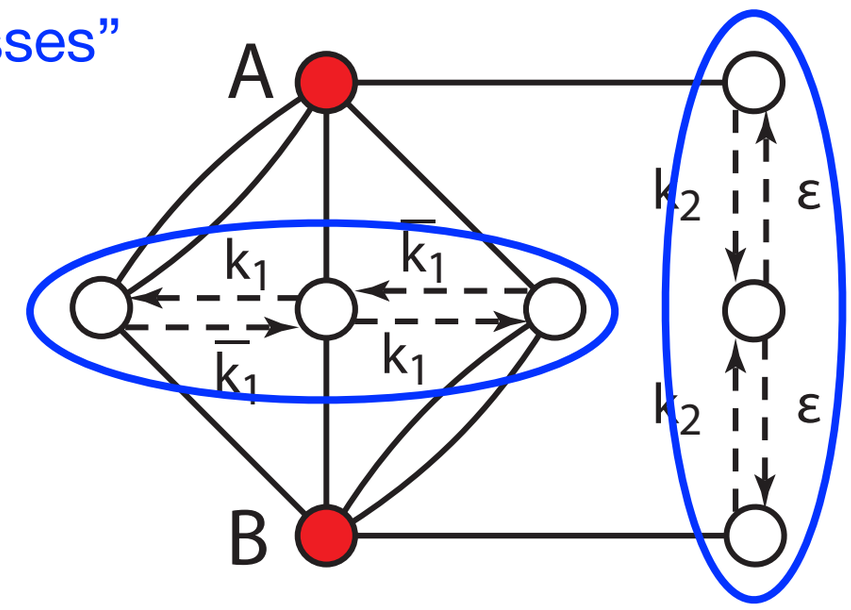




Baez and Biamonte (2017)
Gunawardena (2003)
Feinberg (1979, 1983)
Horn and Jackson (1972)

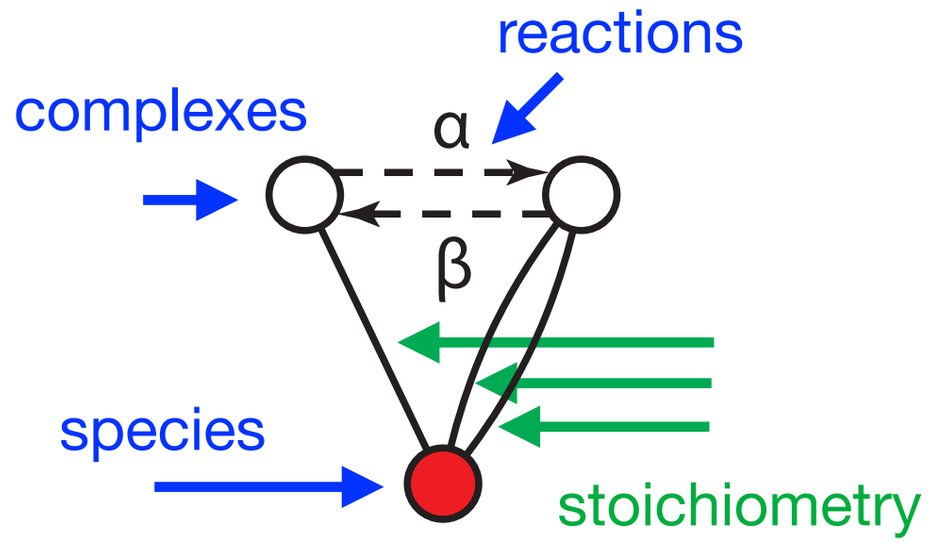
connected components of the
complex graph = “linkage classes”





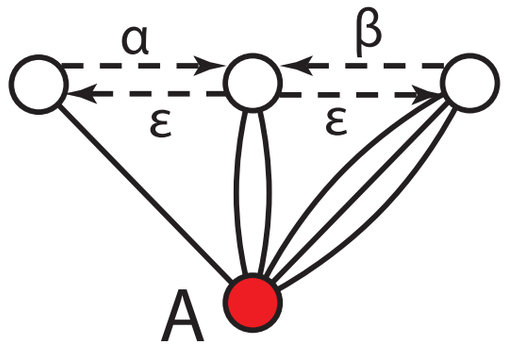
CRN basics: terms and concepts (quantities)

- Species index $p \in 1, \ldots, P$
- Complex index $i, j \in 1, \ldots, \mathcal{C}$
- Reactions (ordered) (i, j) rate const. k_{ji}



- Concentrations (column vector) $n \equiv [n_p]$
- Stoichiometry (complexes to species) $Y \equiv \left[y_p^j\right]$
- Activities (column vector on complexes) $\psi_Y \equiv \left[\psi_Y^i\right] \quad \psi_Y^i \equiv \prod_p n_p^{y_p^i}$
- Adjacency/Rate matrix (btw. complexes) $\mathbb{A} = \sum_{(i,j)} (w_j - w_i)\, k_{ji} w_i^T$
- The rate law $\frac{dn}{dt} = Y \mathbb{A}\, \psi_Y$

Example



rate law

$$\frac{dn}{dt} = Y \mathbb{A}\, \psi_Y$$

$$Y = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

stoichiometric matrix

$$\mathbb{A} = \begin{bmatrix} -\alpha & \epsilon & \\ \alpha & -2\epsilon & \beta \\ & \epsilon & -\beta \end{bmatrix}$$

adjacency matrix

$$\psi_Y = \begin{bmatrix} n \\ n^2 \\ n^3 \end{bmatrix}$$

activities

$$\mathbb{A} = \sum_{(i,j)} (w_j - w_i)\, k_{ji} w_i^T$$

means

$$\mathbb{A} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \alpha \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \epsilon \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} + \cdots$$

The Complex Network: the largest sub-system that is a simple graph

- If this had been a simple graph, the rate law woud have been

$$\frac{d}{dt}\begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix} = \mathbb{A}\begin{bmatrix} n_1 \\ n_2 \\ n_3 \end{bmatrix}$$

$$\mathbb{A} = \begin{bmatrix} -\alpha & \epsilon & \\ \alpha & -2\epsilon & \beta \\ & \epsilon & -\beta \end{bmatrix}$$

- The Key Difference between simple graphs and CRNs:
  - The species always carry the memory of ***state***
  - In graphs, $\mathbb{A}$ acts directly on species
  - In CRNs, $\mathbb{A}$ acts on complexes, which have no state

Part 3: What does topology ensure about dynamics without fine constraints on rates?

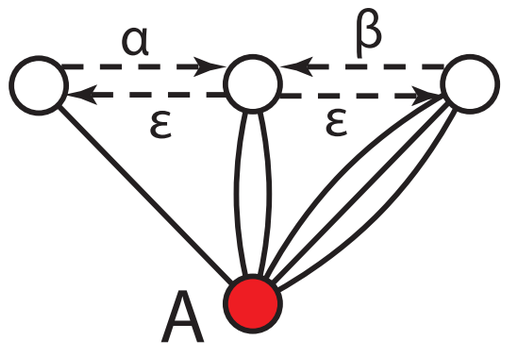
- Two classic results (one now, one in the next section)
  - Feinberg's "deficiency-0" theorem for existence, uniqueness, stability, and properties of classical steady states
  - The Anderson-Craciun-Kurtz theorem for steady-state distributions on deficiency-0 networks

What are the issues?

- In general, the rate equation is ugly: $\frac{dn}{dt} = Y\,\mathbb{A}\,\psi_Y$

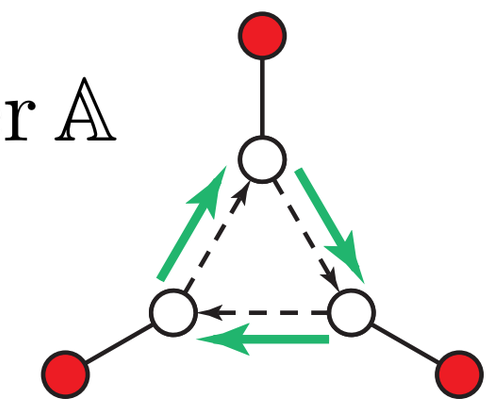
  To what extent does topology allow us to draw conclusions about possible dynamics or steady states?

- Reaction stoichiometry can be “clumpy” — we might expect that, independent of the behavior of the classical rate law, that clumpiness might lead to complicated distributions

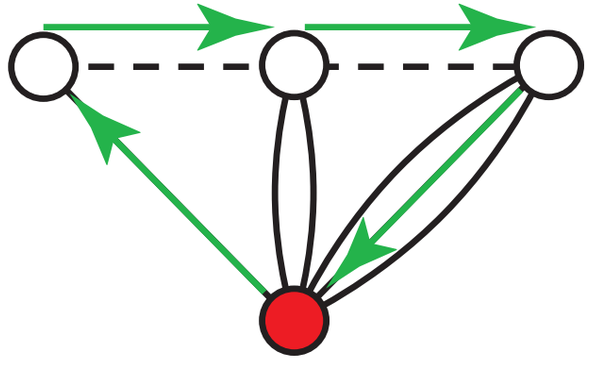


The three types of flows that determine the character of solutions

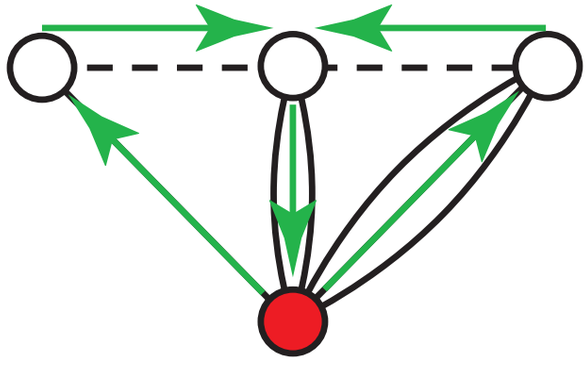
- Balanced flows within the complex network: $\ker \mathbb{A}$



- Flows that change concentration: $\mathrm{Im}(Y\mathbb{A})$



- Balanced flows outside $\ker \mathbb{A}$: $\ker Y \cap \mathrm{Im}\mathbb{A}$



For non-obvious reasons, introduce the concept "deficiency" for a CRN

- Stoichiometric subspace: $S = \mathrm{Im}(Y\mathbb{A}) \quad s \equiv \dim(S)$
- Deficiency: $\delta \equiv \dim(\ker(Y) \cap \mathrm{Im}(\mathbb{A}))$
- Two identities:

$$\begin{aligned}\dim(\mathrm{Im}(\mathbb{A})) &= \dim(\mathrm{Im}(Y\mathbb{A})) + \dim(\ker Y \cap \mathrm{Im}(\mathbb{A})) \\ &= s + \delta \\ C &= \dim(\mathrm{Im}(\mathbb{A})) + \dim(\ker(\mathbb{A})) \\ &= s + \delta + \dim(\ker(\mathbb{A}))\end{aligned}$$

- Lead to a counting rule for deficiency: $\delta = C - s - l$

(complexes) (linkage classes)

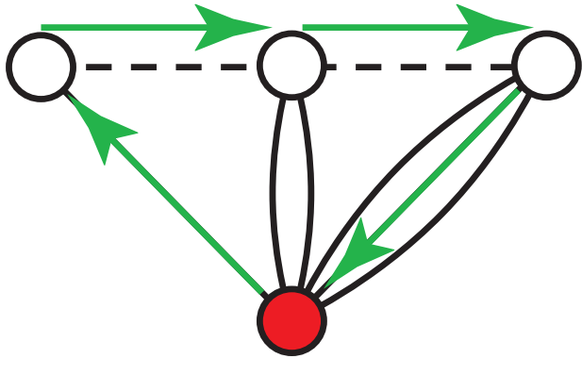
Remarkable result of topology for existence, uniqueness, stability, interiority of steady states

- Theorem (Martin Feinberg)

"If a CRN with δ = 0 is weakly reversible then, for mass action kinetics, the rate equations will have precisely one steady state, within each positive stoichiometric compatibility class. This steady state is asymptotically stable." [and has strictly positive concentrations]

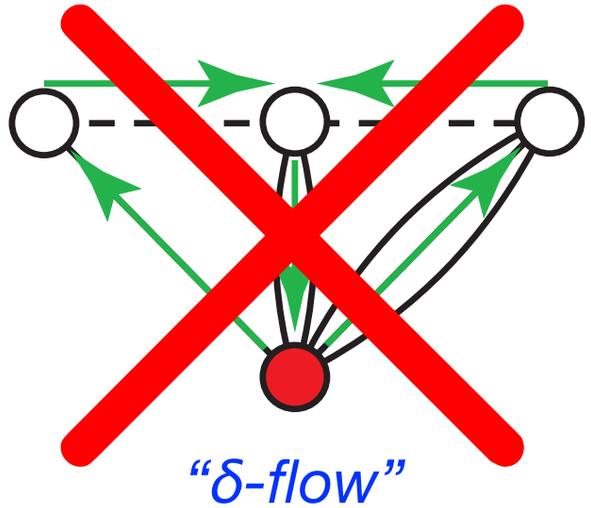
Feinberg and Horn 1974; Horn 1973; Feinberg1977,1980

Intuition for the proof: if deficiency = 0, there can only be mean-regressing flows



“s-flow”

changes some concentration
shifts activities to resist change



“δ-flow”

balanced (integer) hyperflow
preserves concentrations & activities

- If $\delta = 0$, there can be no flows that don’t change concentration of species

- But if species pile up or are depleted, the greater activity drives flows in the opposite direction (interiority)

Part 4: Stochastic CRNs, with a tiny mention of generating-function methods

- Anderson-Craciun-Kurtz Theorem:

"For a CRN with δ = 0, weakly reversible, with mass action kinetics — no matter how clumpy the reactions — the steady-state distributions will be products of Poissons, or slices through them if there are conserved quantities."

D. Anderson, G. Craciun, T. Kurtz 2010

Discrete-state, continuous-time stochastic processes: Master Equation and a taste of ACK

- Probability density function: ρ_{n}
- Master equation: $\frac{d}{dt}\rho_{\mathrm{n}} = \sum_{\mathrm{n}'} \mathrm{T}_{\mathrm{nn}'}\rho_{\mathrm{n}'}$
- Expected numbers: $n \equiv \langle \mathrm{n} \rangle$
- Sampling without replacement leads to mass-action rates

$$\Psi_Y(\mathrm{n}) = \left[\Psi_Y^i(\mathrm{n})\right]$$

$$\Psi_Y^i \equiv \prod_p \frac{\mathrm{n}_p!}{\left(\mathrm{n}_p - y_p^i\right)!}$$

Factorial moments

- A new rate equation: $\frac{d}{dt}\langle \mathrm{n} \rangle = Y\mathbb{A}\langle \Psi_Y(\mathrm{n}) \rangle$

Generating-function methods for discrete-state stochastic processes

- Moment-generating function: $\phi(z) = \sum_{\mathrm{n}} z^{\mathrm{n}} \rho_{\mathrm{n}}$

- If ρ evolves w/ a master equation: $\frac{d}{dt}\rho_{\mathrm{n}} = \sum_{\mathrm{n}'} \mathrm{T}_{\mathrm{nn}'}\rho_{\mathrm{n}'}$

- Then ϕ evolves under a PDE called a Liouville equation: $\frac{\partial}{\partial t}\phi(z) = -\mathcal{L}\left(z, \frac{\partial}{\partial z}\right)\phi(z)$

Factorial moments and Poisson distributions: least-information null models for steady states

- Poisson distribution: $\rho_{\mathrm{n}} = e^{-\xi} \dfrac{\xi^{\mathrm{n}}}{\mathrm{n}!}$

- All factorial moments are simple functions of the mean:

$$\left\langle \frac{\mathrm{n}!}{(\mathrm{n}-k)!} \right\rangle = \sum_{\mathrm{n}} \rho_{\mathrm{n}} \frac{\mathrm{n}!}{(\mathrm{n}-k)!} = \xi^k$$

- The generating functions are exponentials:

$$\phi_\xi(z) \equiv \sum_{\mathrm{n}} e^{-\xi} \frac{z^{\mathrm{n}} \xi^{\mathrm{n}}}{\mathrm{n}!} = e^{(z-1)\xi}$$

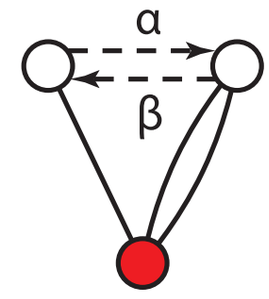
- They are eigenstates of the derivative operator:

$$\frac{\partial}{\partial z} \phi_\xi(z) = \xi \phi_\xi(z)$$

Example at δ = 0, and elegant proof that a Poisson distribution is a steady-state solution

- Master equation:

$$\frac{d\rho_{\mathrm{n}}}{dt} = \left\{ \left(e^{-\partial/\partial \mathrm{n}} - 1 \right) \alpha \mathrm{n} + \left(e^{\partial/\partial n} - 1 \right) \beta \mathrm{n} \left(\mathrm{n} - 1 \right) \right\} \rho_{\mathrm{n}}$$



- Liouville function:

$$\mathcal{L} = \begin{bmatrix} z & z^2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} \alpha & -\beta \end{bmatrix} \begin{bmatrix} \partial/\partial z \\ \partial^2/\partial z^2 \end{bmatrix}$$

(rate matrix)

$$= z(1-z) \frac{\partial}{\partial z} \left(\alpha - \beta \frac{\partial}{\partial z} \right)$$

- But on a Poisson distribution: $\left(\alpha - \beta \dfrac{\partial}{\partial z} \right) \phi_\xi(z) = (\alpha - \beta\xi)\, \phi_\xi(z)$

- So steady-state condition $\mathcal{L}\phi_\xi(z) = 0$ whenever $\xi = \alpha/\beta$

Getting general ACK from Feinberg's deficiency-0

- More general form for $\mathcal{L}$: $\mathcal{L} = -\psi_Y^T(z)\,\mathbb{A}\,\psi_Y\left(\frac{\partial}{\partial z}\right)$

General behavior on a gen. fn. $\psi_Y\left(\frac{\partial}{\partial z}\right)\phi_\xi(z)\Big|_{z=1} = \langle \Psi_Y(\mathrm{n})\rangle$

For a Poisson, remember factorial moments: $= \psi_Y(\langle \mathrm{n}\rangle)$

The derivative just pulls out the coherent-state parm. $= \psi_Y(\xi)$

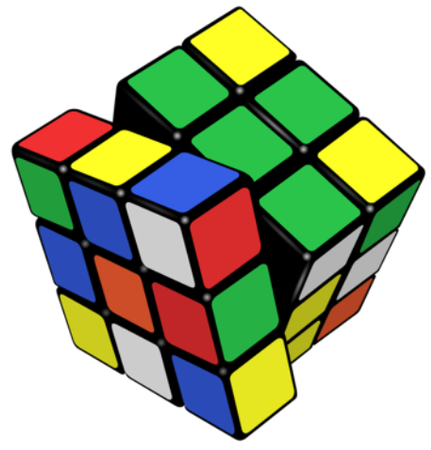
$$\mathcal{L}\phi_\xi(z) = -\psi_Y^T(z)\,\mathbb{A}\,\psi_Y(\xi)\phi_\xi(z)$$

- This is just the condition in the rate equation that Feinberg shows can always be solved at δ = 0. END PROOF

$$\frac{dn}{dt} = Y\,\mathbb{A}\,\psi_Y(n)$$

More that can be done

- Beyond deficiency-0? When (else) are CRNs simple?
- Duality and fluctuation theorems (Jarzynski, Crooks)
- Develop stat. mech. of concurrency further:
  - Systems-biology language Kappa (Fontana, Danos)
  - Large-Deviations for stochastic Rubik's Cube?



- Category theory: rule & network aggregation / reduction

Concluding remarks

- Diffusion on ordinary graphs is a restrictive framework, but has still be extremely useful

- Adding concurrency and clumpiness with CRNs is a natural next step, but adds considerable mathematical difficulty

- This ties in well with macro-worlds, and the area of overlap contains a lot of interesting work to do