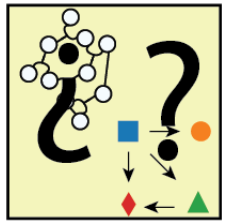
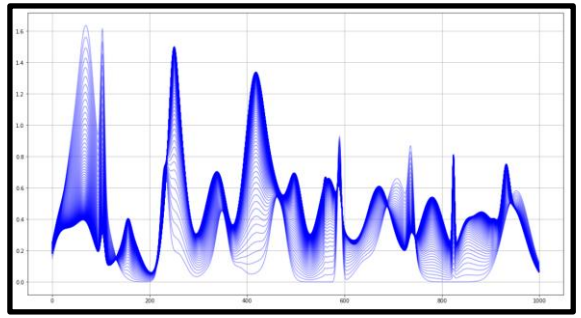
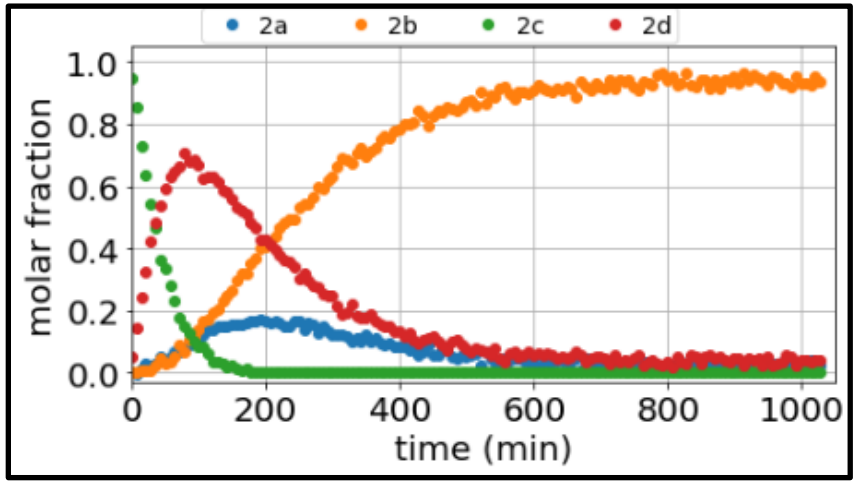
Structural signatures of CRNs in reaction-monitoring data

Alex Blokhuis

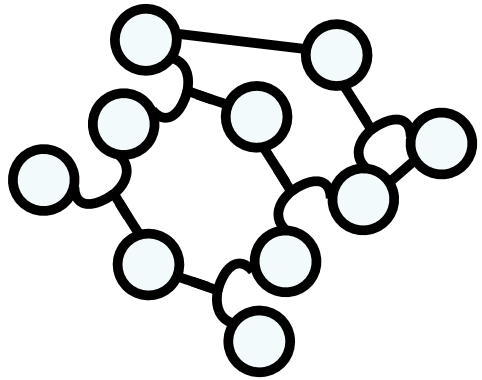
Problem specification: data → chemical reaction network (CRN)







Observations, data

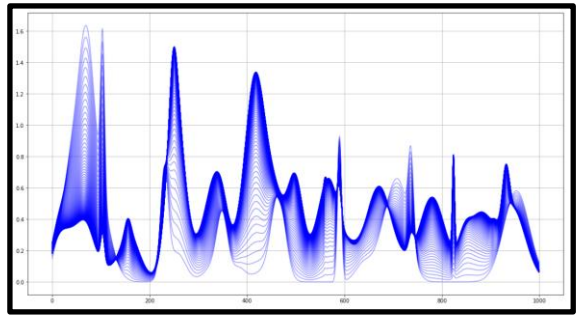


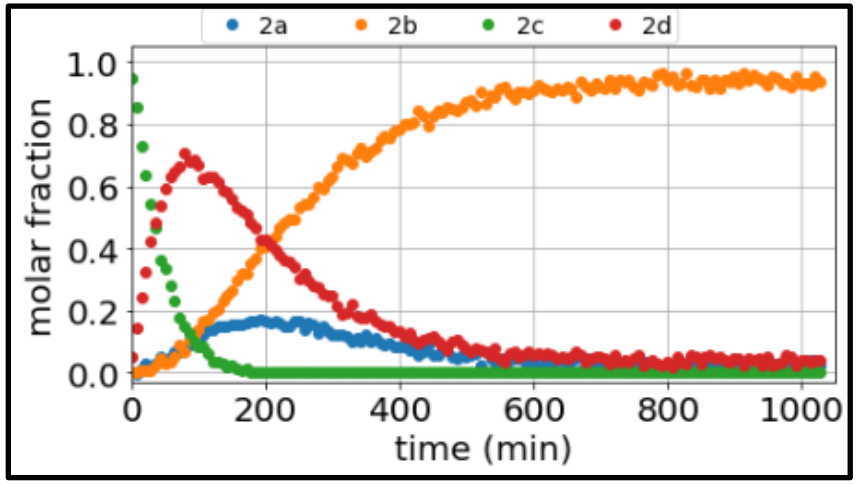
$$\mathrm{ArSII} + \mathrm{Cl_2} \overset{\mathrm{I}}{\leftrightarrows} \mathrm{IICl} + \mathrm{ArSCl}$$
$$\mathrm{NCS} + \mathrm{IICl} \overset{\mathrm{II}}{\leftrightarrows} \mathrm{NCII} + \mathrm{Cl_2}$$
$$\mathrm{NCS} + \mathrm{ArII} \overset{5}{\leftrightarrows} \mathrm{NIIS} + \mathrm{ArSCl}$$
$$\mathrm{ArSCl} + \mathrm{ArSII} \overset{6}{\leftrightarrows} \mathrm{IICl} + \mathrm{AR_2S_2}$$
$$\mathrm{Ar_2S_2} + \mathrm{Cl_2} \overset{7}{\leftrightarrows} 2\ \mathrm{ArSCl}$$

CRN:

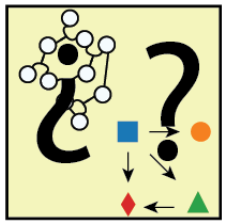
Species + reactions

A question of methodology:
How to extract structure from data?





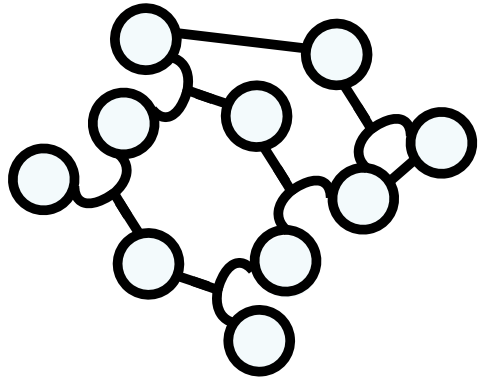
Observations, data



Today, this proces is (in general):

i) Not systematic, lacks methodology

ii) Arduous and time-intensive



$$\begin{aligned}
&\mathrm{ArSH} + \mathrm{Cl_2} \overset{\mathrm{I}}{\leftrightarrows} \mathrm{HCl} + \mathrm{ArSCl} \\
&\mathrm{NCS} + \mathrm{HCl} \overset{\mathrm{II}}{\leftrightarrows} \mathrm{NCH} + \mathrm{Cl_2} \\
&\mathrm{NCS} + \mathrm{ArH} \overset{5}{\leftrightarrows} \mathrm{NHS} + \mathrm{ArSCl} \\
&\mathrm{ArSCl} + \mathrm{ArSH} \overset{6}{\leftrightarrows} \mathrm{HCl} + \mathrm{AR_2S_2} \\
&\mathrm{Ar_2S_2} + \mathrm{Cl_2} \overset{7}{\leftrightarrows} 2\ \mathrm{ArSCl}
\end{aligned}$$

CRN:

Species + reactions

Methodology: Descartes

1. **Filter:** Accept only what is true beyond reasonable doubt

2. **Division:** Split problems up in smallest parts

3. **Solve:** Simple problems first

4. **Exhaustion:** Generalize & enumerate, cover all possibilities

Methodology: Descartes

1. Filter: Accept only what is true beyond reasonable doubt

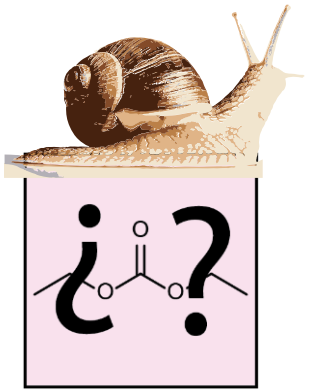
2. Division: Split problems up in smallest parts

3. Solve: Simple problems first

4. Exhaustion: Generalize & enumerate, cover all possibilities

Chemistry has solved questions of structure before, let's see what we did then.

Analytical Challenges – elucidating structure (molecules)

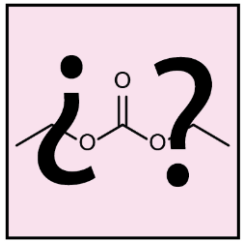


Organic chemistry (<1960s) used to center around identifying compounds. This process eventually became systematized, involving many steps of examination and experimental tests.

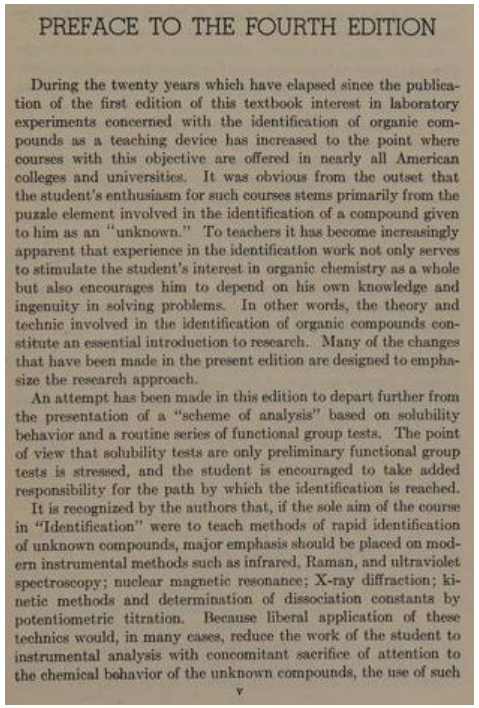
- **Preliminary Examination:** homogeneity, state, color, odor, ignition test
- **Physical constants:** Melting point, boiling point
- **Elemental analysis**
- **Solubility tests** (in H_2O, dil. HCl, dil. NaOH, NaHCO3, cold H2SO4, H3PO4, ether).
- **Classification tests** (for functional groups, unsaturation, halogens, acids, alcohols, amines, aldehydes, ketones, aromatics, ethers, esters, nitro, phenol)
- **Literature comparison**
- **Preparation of derivatives** (+ analysis thereof, e.g. specific gravity, refractive index, melting point, optical rotation, . . .)
- **If molecule is new: fragmentation + characterization of fragments**

(See also The Systematic Identification Of Organic Compounds 3rd edition (1940))

Analytical Challenges – elucidating structure (molecules)



Organic chemistry (>1960s) was dramatically transformed by analytical techniques, **allowing to focus on myriad other topics than identification**



PREFACE TO THE FOURTH EDITION

During the twenty years which have elapsed since the publication of the first edition of this textbook interest in laboratory experiments concerned with the identification of organic compounds as a teaching device has increased to the point where courses with this objective are offered in nearly all American colleges and universities. It was obvious from the outset that the student's enthusiasm for such courses stems primarily from the puzzle element involved in the identification of a compound given to him as an "unknown." To teachers it has become increasingly apparent that experience in the identification work not only serves to stimulate the student's interest in organic chemistry as a whole but also encourages him to depend on his own knowledge and ingenuity in solving problems. In other words, the theory and technic involved in the identification of organic compounds constitute an essential introduction to research. Many of the changes that have been made in the present edition are designed to emphasize the research approach.

An attempt has been made in this edition to depart further from the presentation of a "scheme of analysis" based on solubility behavior and a routine series of functional group tests. The point of view that solubility tests are only preliminary functional group tests is stressed, and the student is encouraged to take added responsibility for the path by which the identification is reached.

It is recognized by the authors that, if the sole aim of the course in "Identification" were to teach methods of rapid identification of unknown compounds, major emphasis should be placed on modern instrumental methods such as infrared, Raman, and ultraviolet spectroscopy; nuclear magnetic resonance; X-ray diffraction; kinetic methods and determination of dissociation constants by potentiometric titration. Because liberal application of these technics would, in many cases, reduce the work of the student to instrumental analysis with concomitant sacrifice of attention to the chemical behavior of the unknown compounds, the use of such

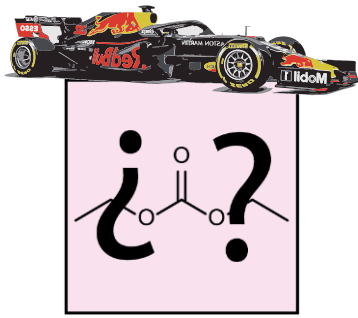
v

"If the sole aim of the course in "identification" were to teach methods of rapid identification of unknown compounds, major emphasis should be placed on modern instrumental methods such as infrared, Raman, and ultraviolet spectroscopy; nuclear magnetic resonance; X-ray diffraction; kinetic methods and determination of dissociation constants by potentiometric titration."

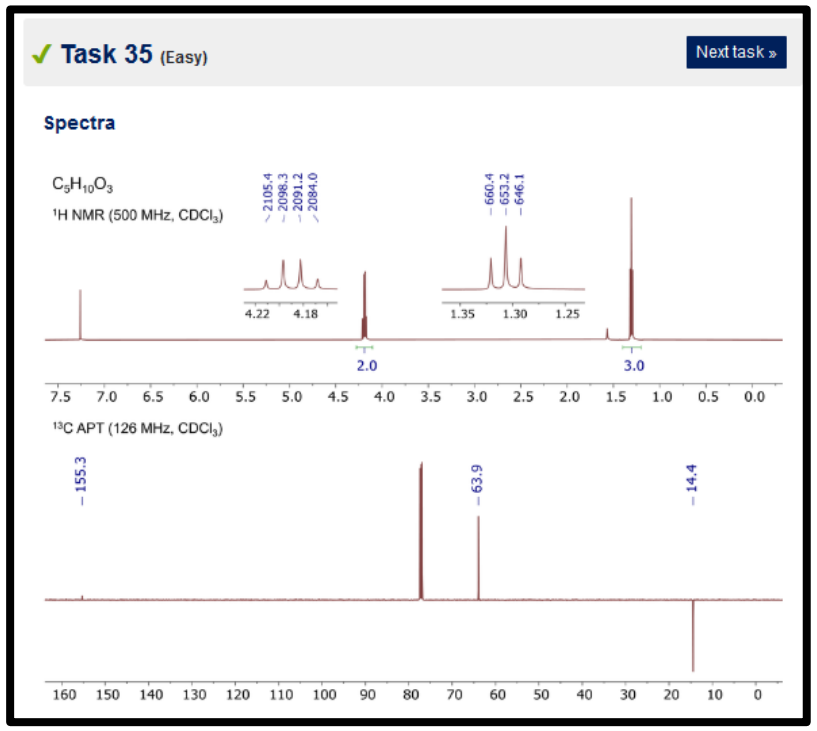
"Because liberal application of these techniques would, in many cases, reduce the work of the student to instrumental analysis with concomitant sacrifice of attention to the chemical behavior of the unknown compounds, the use of such technics has been strictly limited"

(from The Systematic Identification Of Organic Compounds 4th edition (1956))

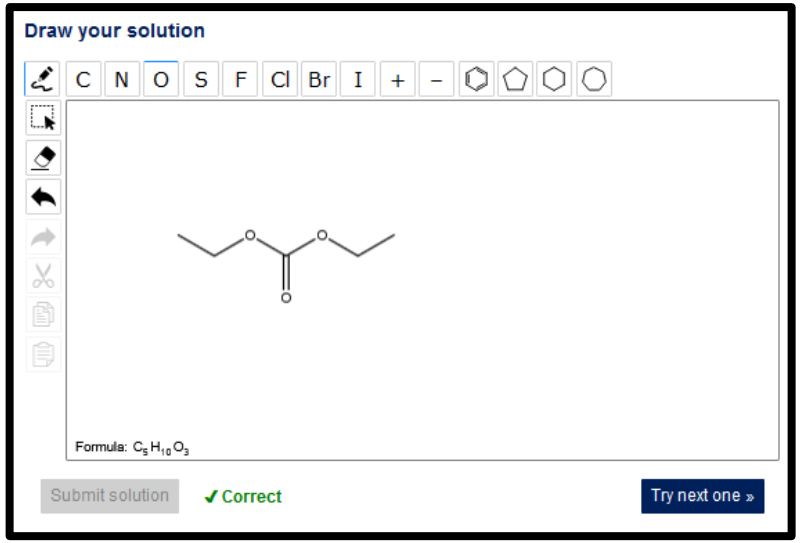
Analytical Challenges – elucidating structure (molecules)



Today, elucidating structures of (small) molecules has become
a quick puzzle you can do for fun on the internet



NMR CHALLENGE
IOCB Prague

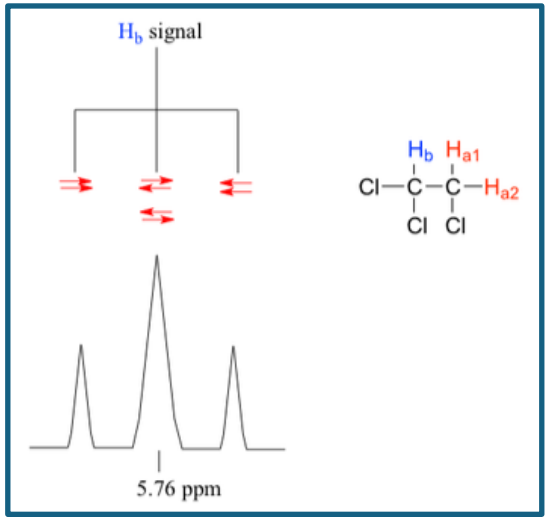


Why is NMR so efficient?

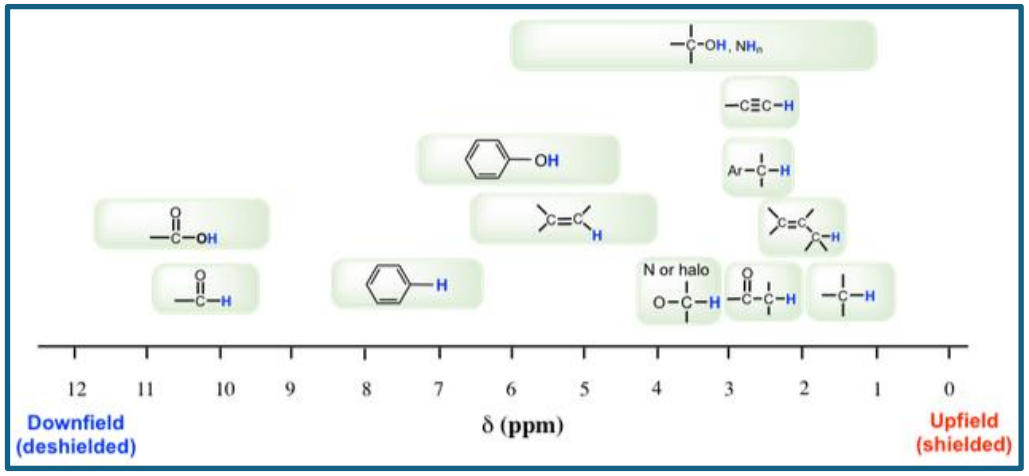
All elucidation **combines structural clues to filter hypotheses**

NMR:

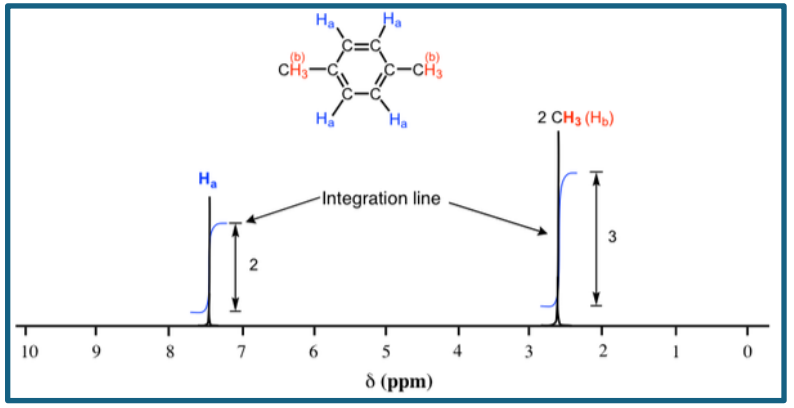
- Measurement of **local structure** through **interpretable** indices
- **Scaling:** more complex structures give more distinct clues



Coupling + Splitting :
near neighbors, # near nuclei



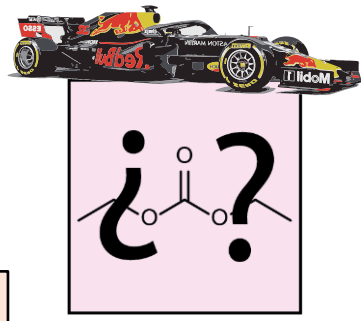
Chemical shift (δ):
Functional groups

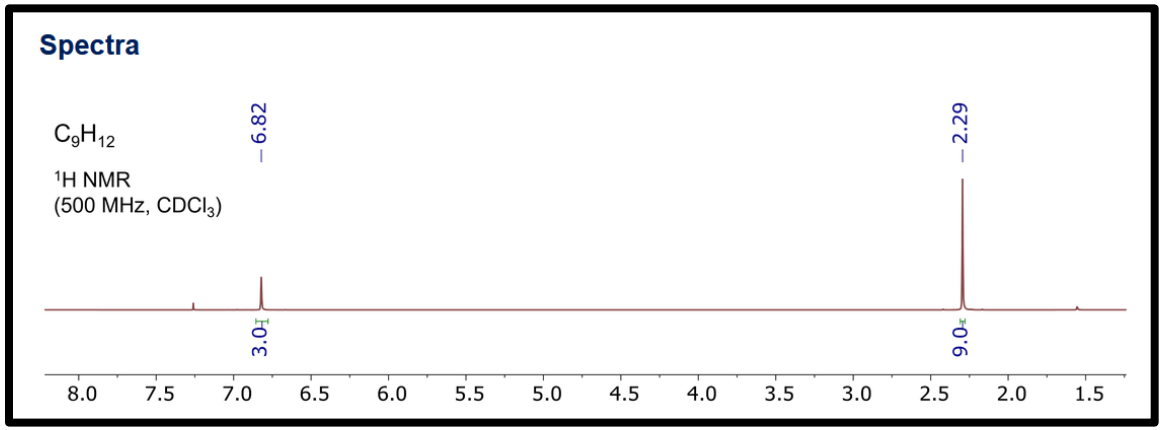


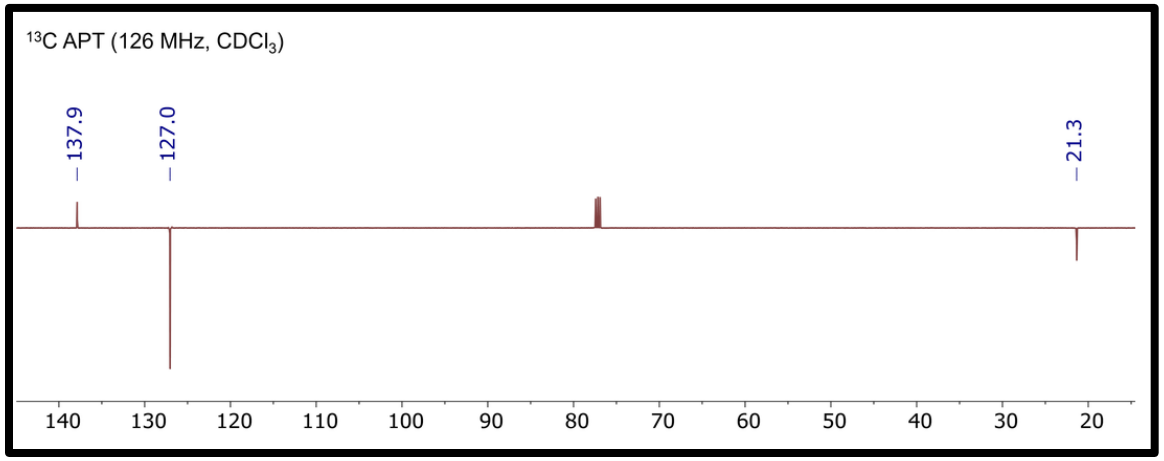
Integrals:
(rel.) # local nuclei

And more . . .

NMR as a puzzle

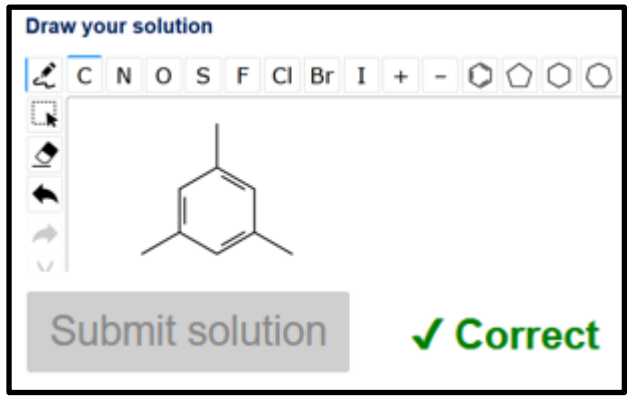




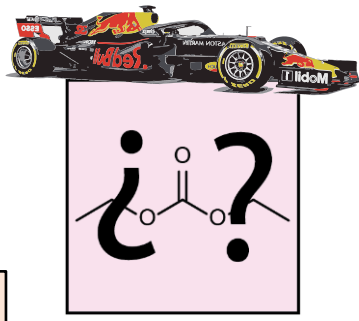


C_9H_{12} (>4000 isomers)

Nmr indices:
HNMR peaks,
CNMR peaks,
peak integrals,
peak multiplicity,
peak couplings,
APT signs, . . .

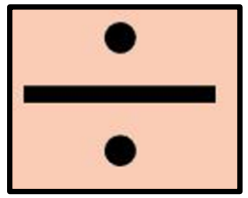


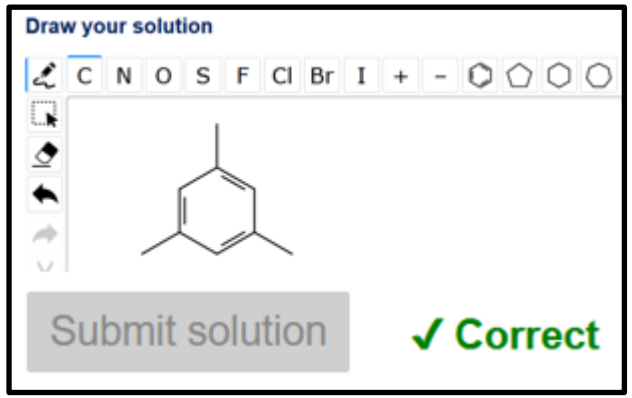
Indices exponentially reduce candidate structures



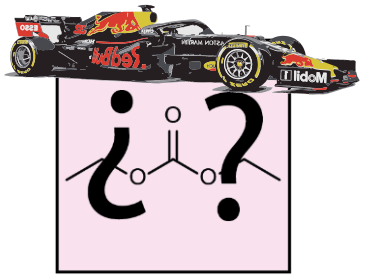
C_9H_{12} (>4000 isomers)

Nmr indices:
HNMR peaks,
CNMR peaks,
peak integrals,
peak multiplicity,
peak couplings,
APT signs, . . .



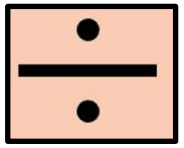


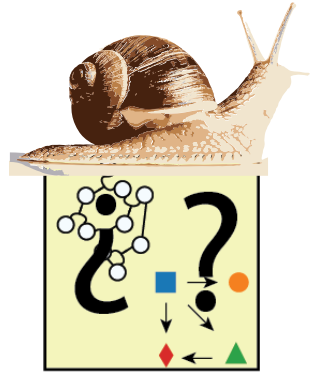
Analytical Challenges – elucidating structure (CRNs)



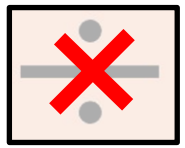
Today, we can quickly elucidate (small) molecular structures.

(Through **scalable, interpretable, structural** indices)

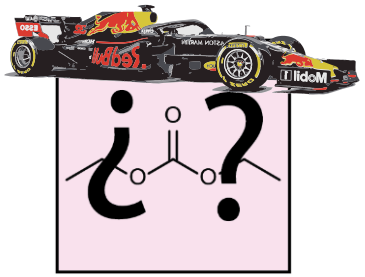




Elucidating chemical reaction networks (CRNs) is still **hard** and **slow**.

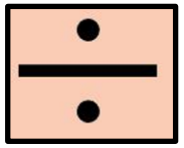


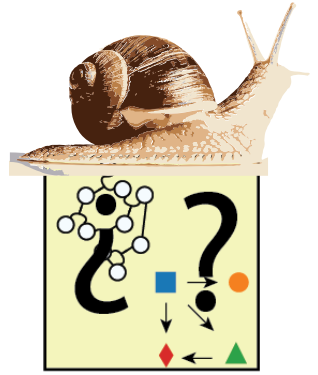
Analytical Challenges – elucidating structure (CRNs)



Today, we can quickly elucidate (small) molecular structures.

(Through **scalable, interpretable, structural** indices)

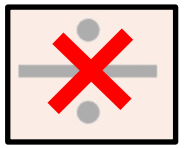




Elucidating chemical reaction networks (CRNs) is still **hard** and **slow**.

There is no "*Systematic Identification Of CRNs*"

But we do already have analytical techniques for Reaction Monitoring



Analytical Challenges – elucidating structure (CRNs)

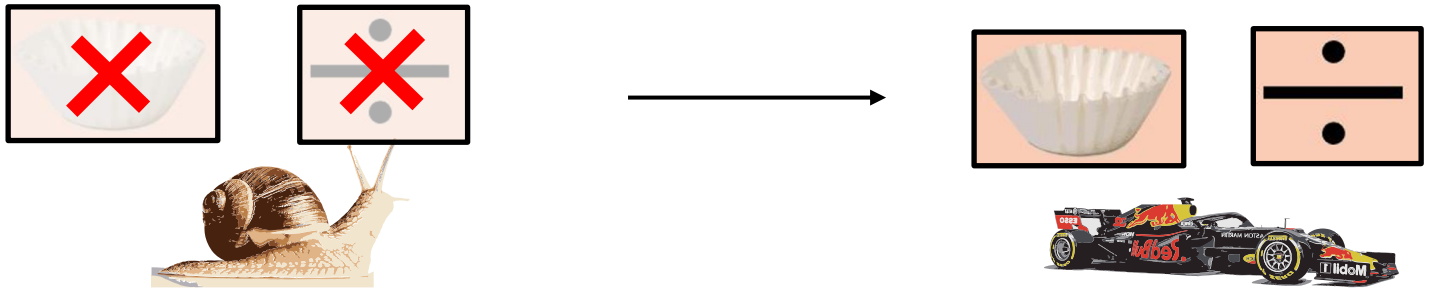
Today, we can quickly elucidate (small) molecular structures.

(Through **scalable, interpretable, structural** indices)

Elucidating chemical reaction networks (CRNs) is still **hard** and **slow**.

There is no "*Systematic Identification Of CRNs*"

But we do already have analytical techniques for Reaction Monitoring
what is still missing are **scalable, interpretable, structural** indices.



Analytical Challenges – elucidating structure (CRNs)

Today, we can quickly elucidate (small) molecular structures.

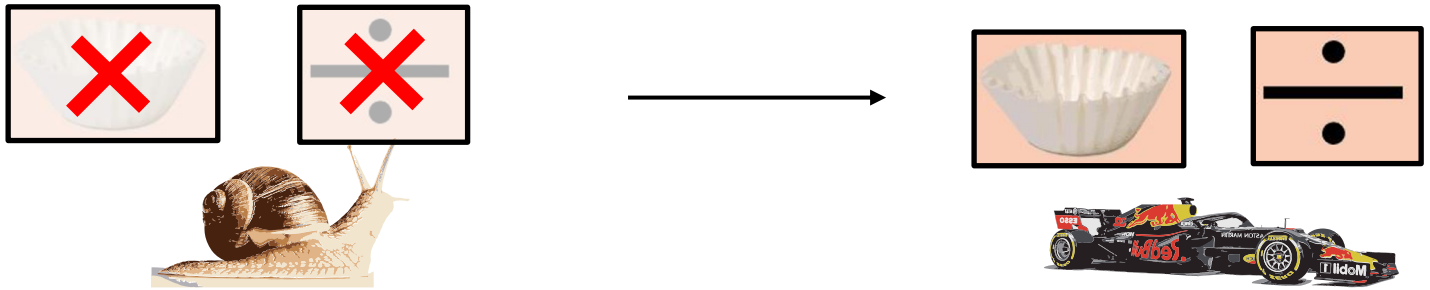
(Through **scalable, interpretable, structural** indices)

Elucidating chemical reaction networks (CRNs) is still **hard** and **slow**.

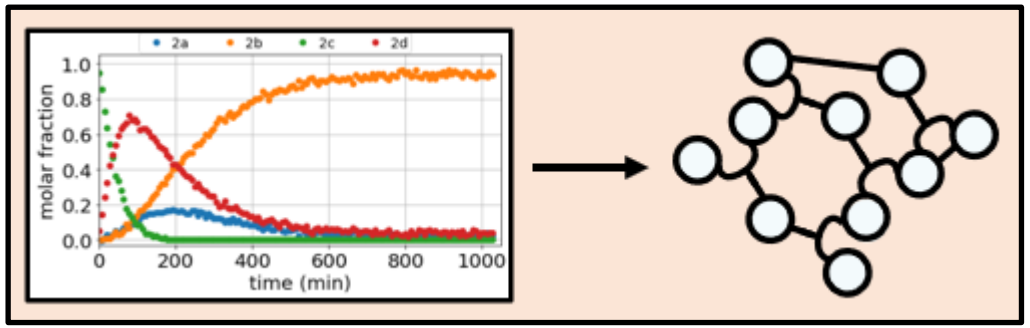
There is no "*Systematic Identification Of CRNs*"

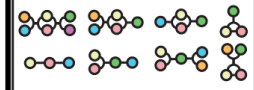
But we do already have analytical techniques for Reaction Monitoring
what is still missing are **scalable, interpretable, structural** indices.

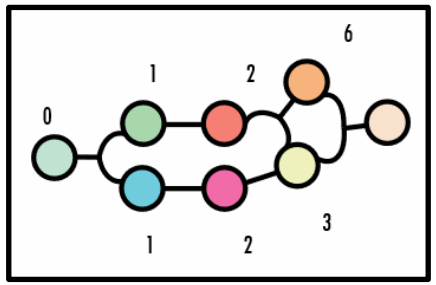
i.e. unambiguous clues about CRN structure extractable from data

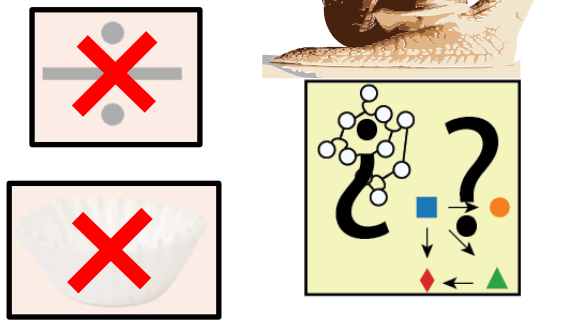


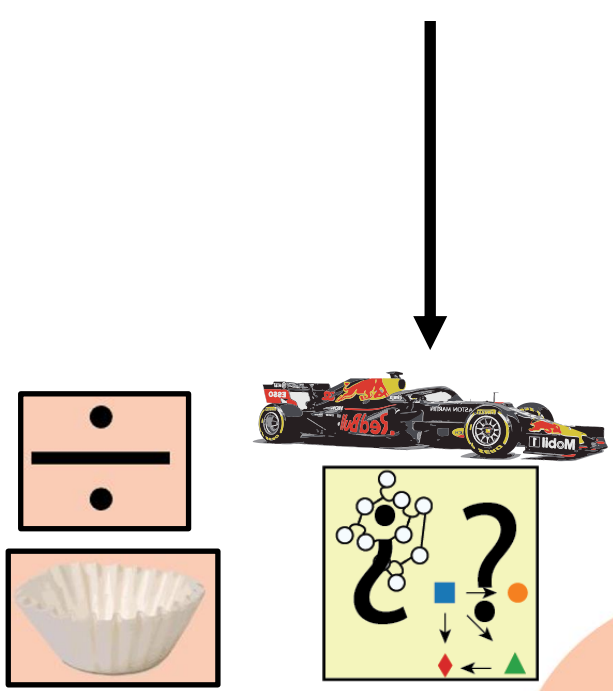
My research line: structural indices of CRNs from reaction-monitoring data



CRN	d	d^+	$d^\wedge$	I^x
	2	3	3	0
	2	3	4	1
	2	4	4	0
	2	4	4	1
	2	4	4	2
	2	4	5	1
	2	4	5	2
	2	4	6	3

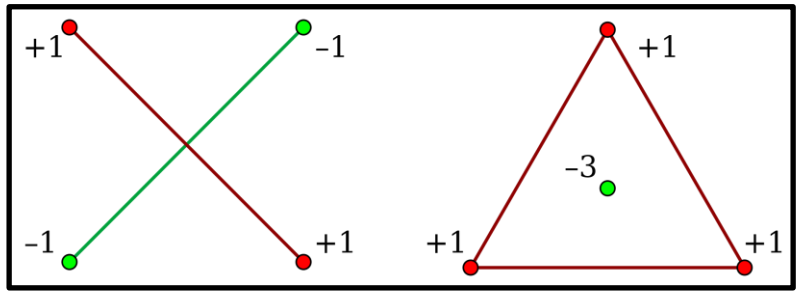




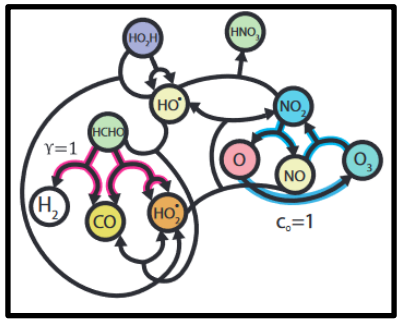


Some measurable structural indices

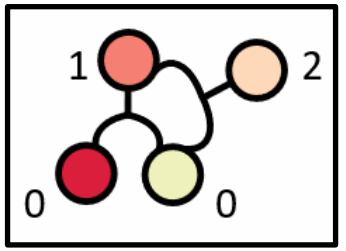
Discussed in Doubice, see slides in Discord

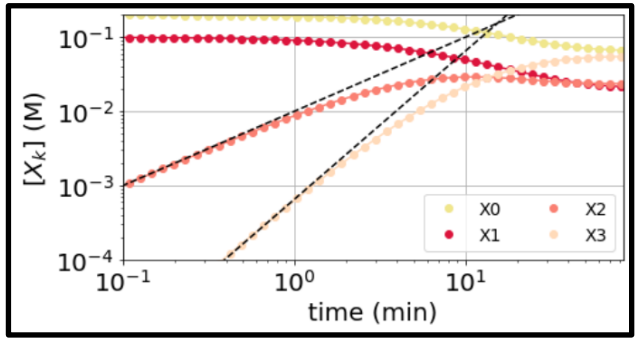


Radon partitions (generalized notions of convexity)



Data dimension, Nullspace





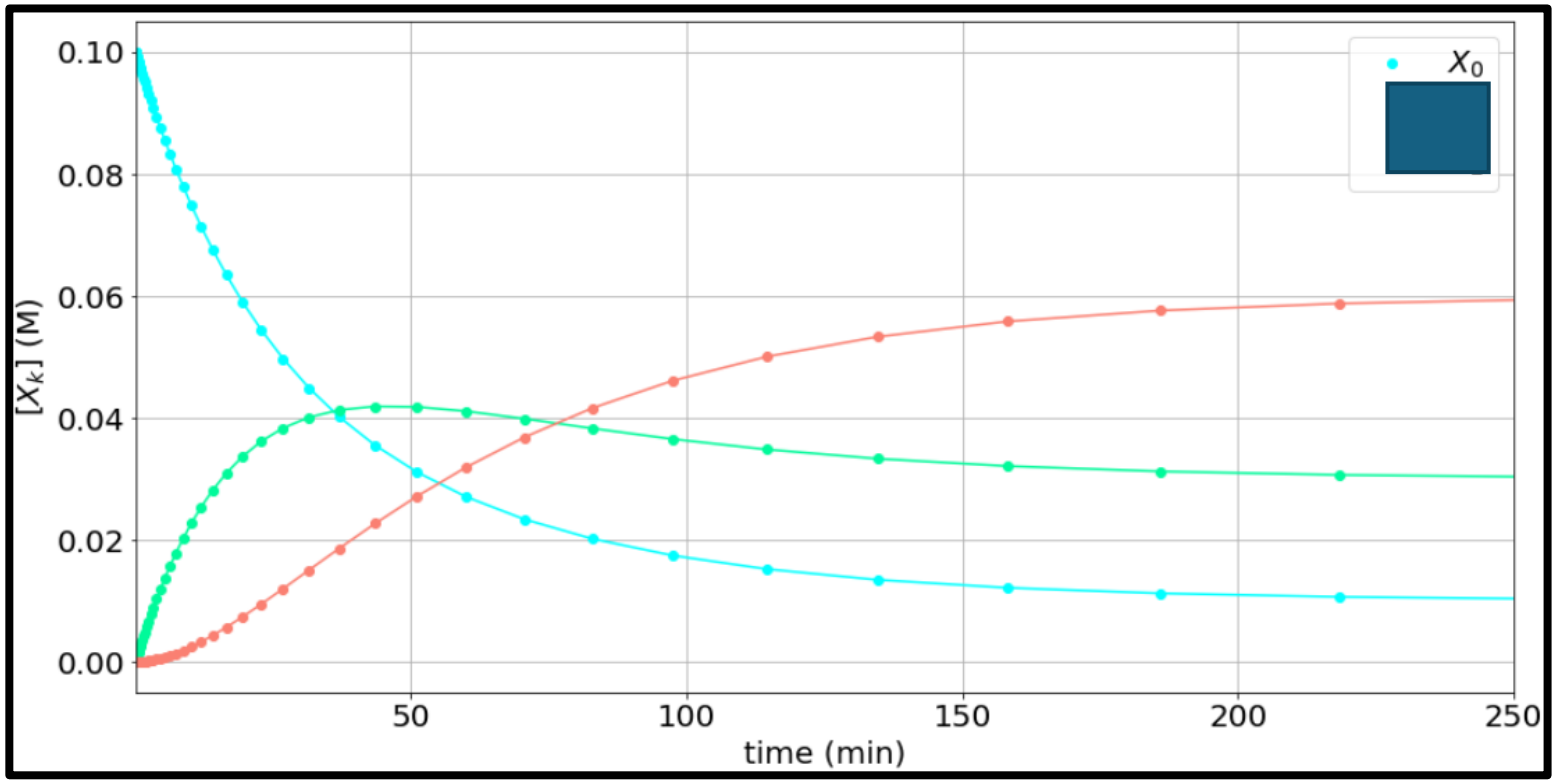
Kinetic exponents*

$$[\mathrm{X_k}] = a_0 + a_1\, t + a_2\, t^2 + \ldots$$

*similar to Delplot Rank

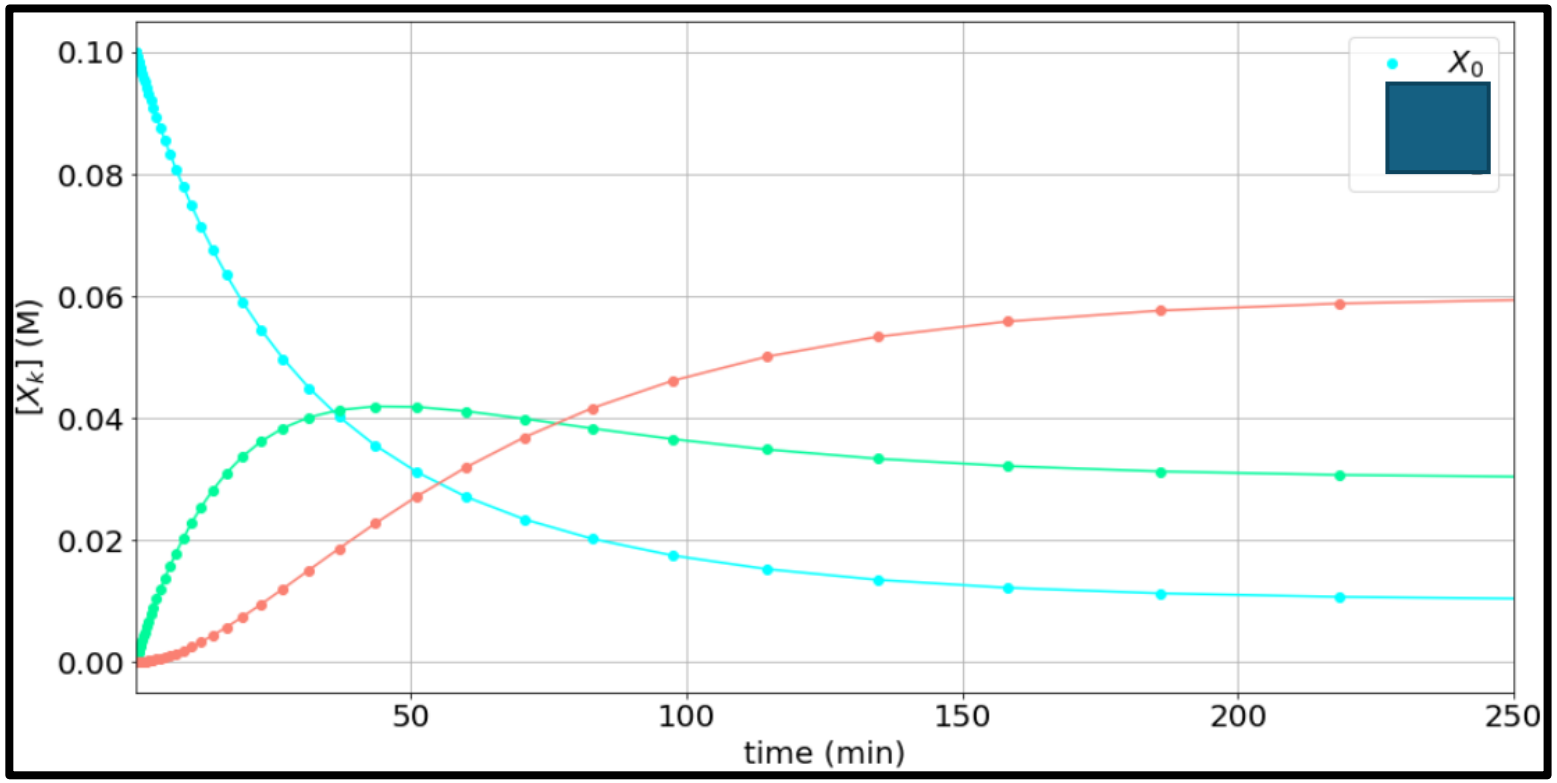
KENA: 2 reaction steps

$$X_0 \rightleftarrows X_1 \rightleftarrows X_2$$



2 reaction steps

$$X_0 \rightleftarrows X_1 \rightleftarrows X_2$$



Q1: which is X_1, which is X_2?

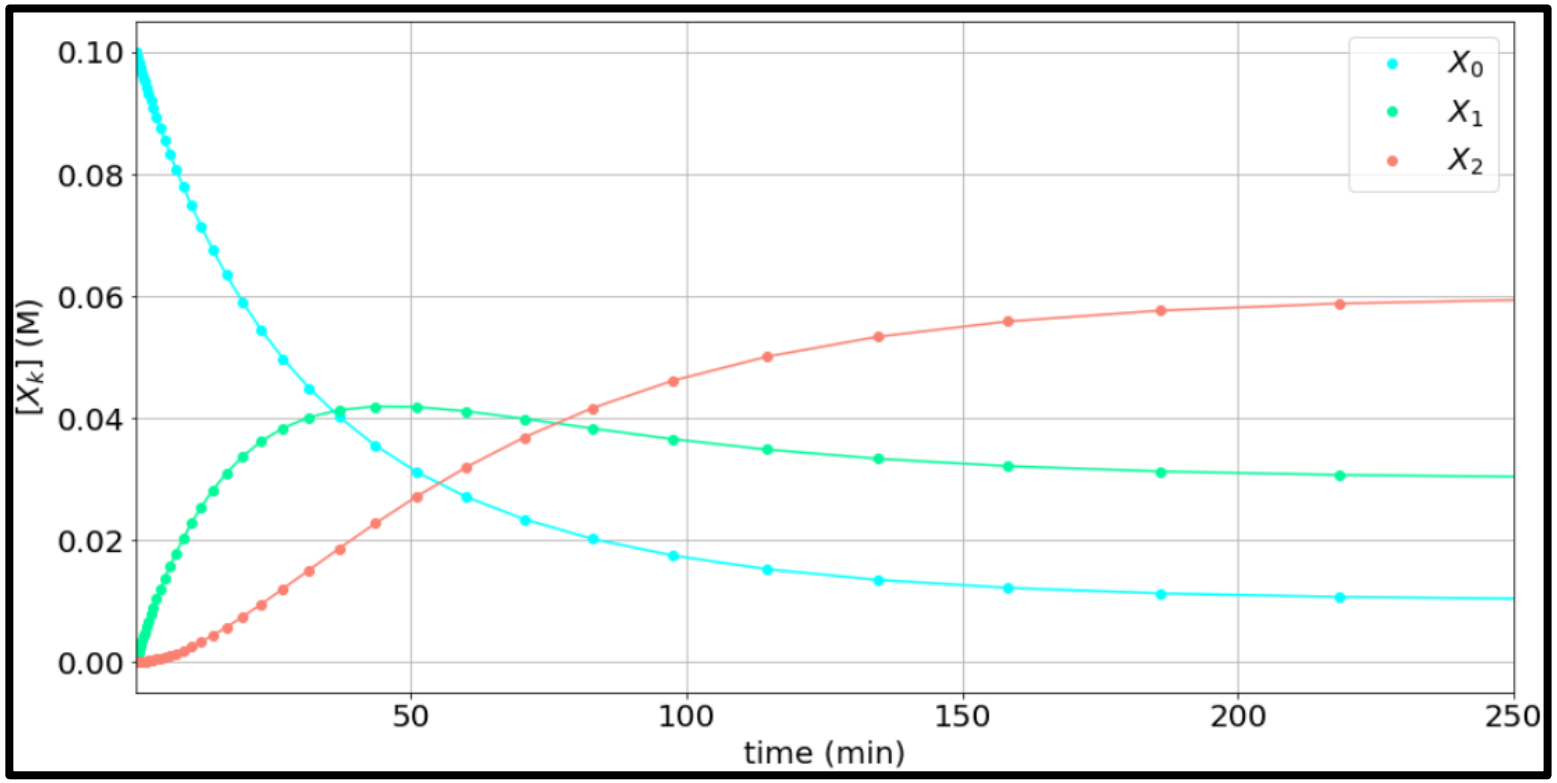
Q2: can we show it without

Trial & error / data fitting?

i.e. can we prove it, or **measure** it, 'nmr-style'?

2 reaction steps

$$X_0 \rightleftarrows X_1 \rightleftarrows X_2$$



Q1: which is X_1, which is X_2?

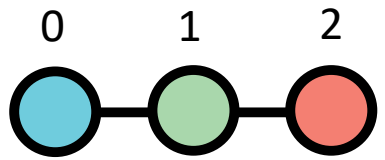
Q2: can we show it without

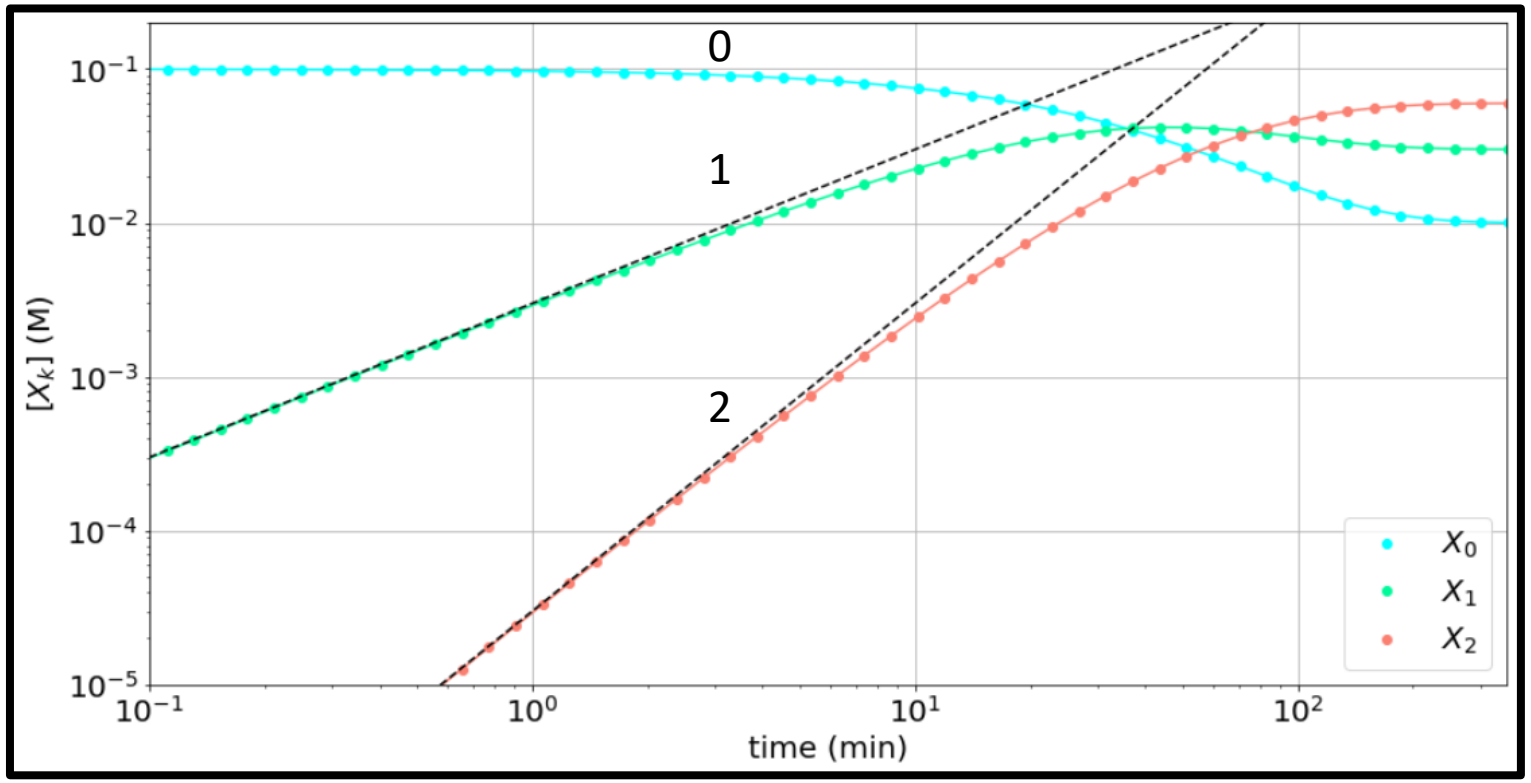
Trial & error / data fitting?

i.e. can we **measure** it?

Log-log

$$X_0 \rightleftarrows X_1 \rightleftarrows X_2$$





Q1: which is X_1, which is X_2?

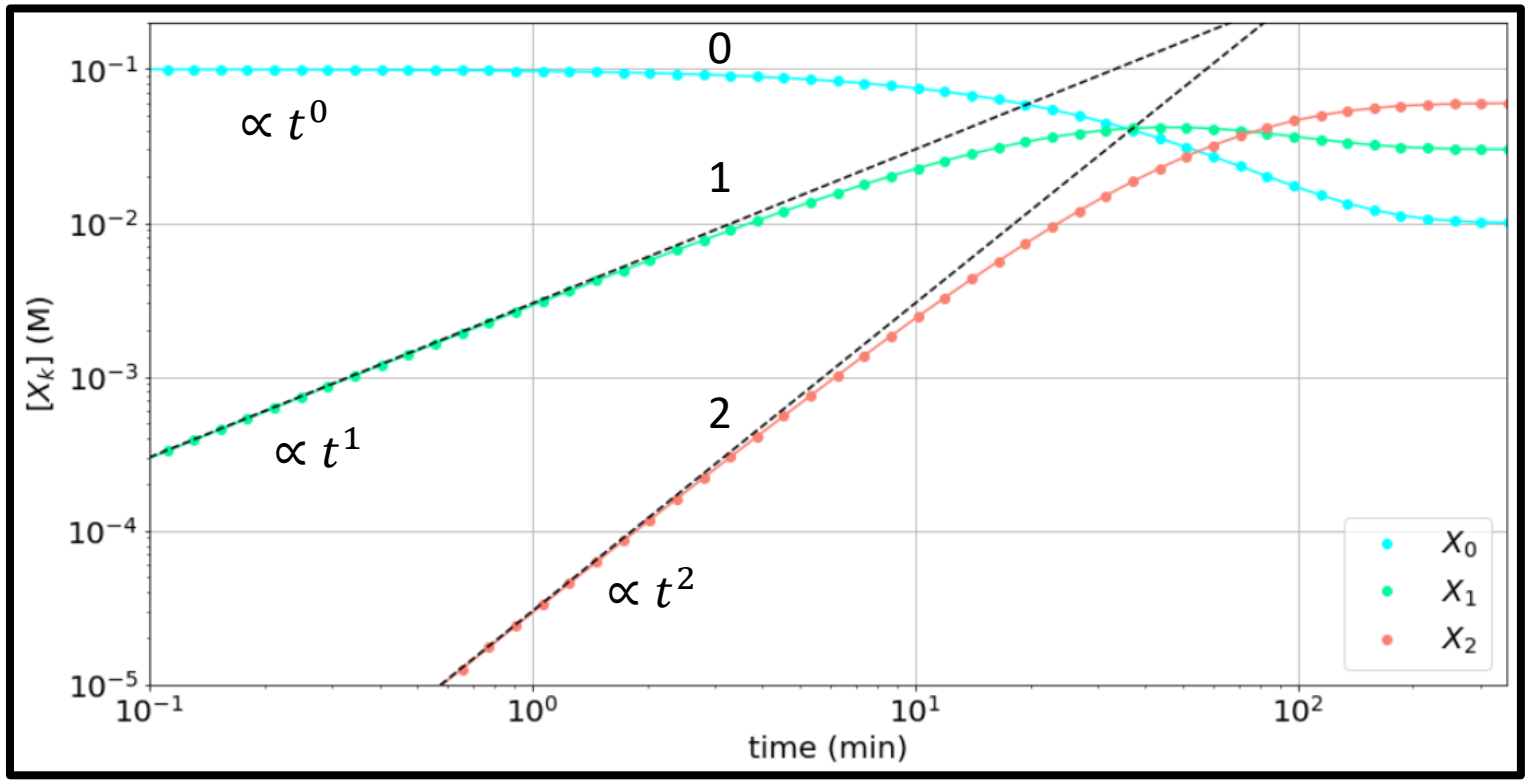
Q2: can we show it without

Trial & error / data fitting?

i.e. can we **measure** it?

Log-log reveals “Kinetic exponents”

$$X_0 \rightleftarrows X_1 \rightleftarrows X_2$$



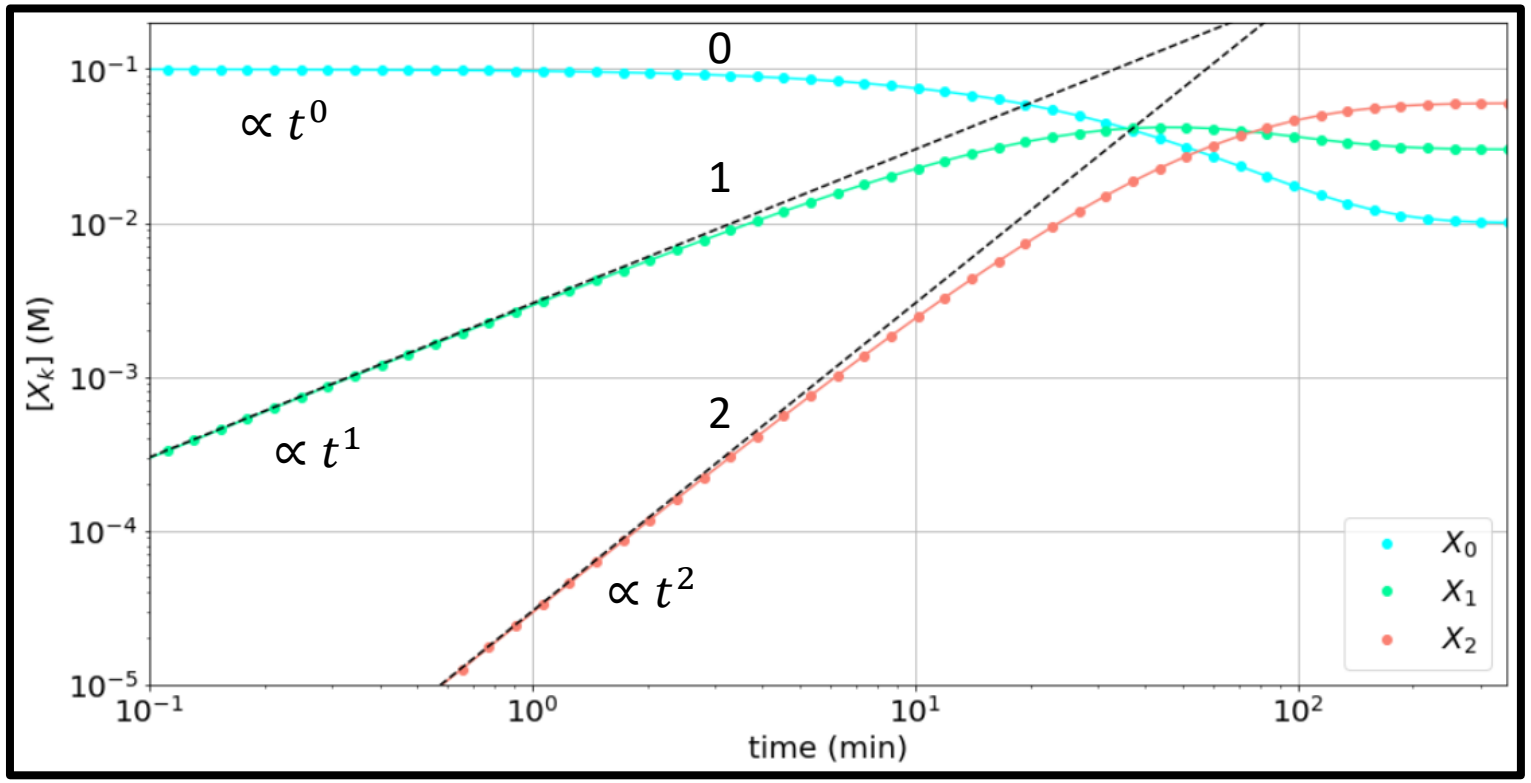
$$[X_0](0^+) \propto t^0$$
$$[X_1](0^+) \propto t^1$$
$$[X_2](0^+) \propto t^2$$

Leading terms in dynamics as function of time contain network connectivity

(proof in a few slides)

Log-log reveals “Kinetic exponents”

$$X_0 \rightleftarrows X_1 \rightleftarrows X_2$$



$$[X_0](0^+) \propto t^0$$
$$[X_1](0^+) \propto t^1$$
$$[X_2](0^+) \propto t^2$$

Leading terms in dynamics as function of time contain network connectivity

(proof in a few slides)

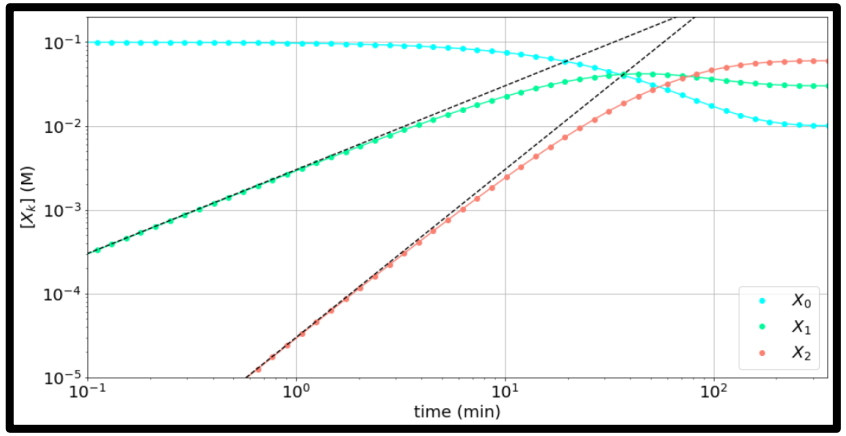
Theory I – linear CRNs

$$d_t[\mathbf{X}] = \mathbb{J}\,[\mathbf{X}]$$

General solution:

$$[\mathbf{X}] = \exp(\mathbb{J}\,t)[\mathbf{X}]_0$$

$$[\mathrm{X_k}] = [\mathrm{X_k}]_0 + t\,(\mathbb{J}\,[\mathbf{X}]_0) + t^2\frac{(\mathbb{J}^2[\mathbf{X}]_0)}{2} + t^3\frac{(\mathbb{J}^3[\mathbf{X}]_0)}{3!} + \ldots$$



Theory I – linear CRNs

$$d_t[\mathbf{X}] = \mathbb{J}\,[\mathbf{X}]$$

General solution:

$$[\mathbf{X}] = \exp(\mathbb{J}\,t)[\mathbf{X}]_0$$

$$[\mathrm{X_k}] = [\mathrm{X_k}]_0 + t\,(\mathbb{J}\,[\mathbf{X}]_0) + t^2\frac{(\mathbb{J}^2[\mathbf{X}]_0)}{2} + t^3\frac{(\mathbb{J}^3[\mathbf{X}]_0)}{3!} + \ldots$$

Path interpretation:

Change due to 1 reaction

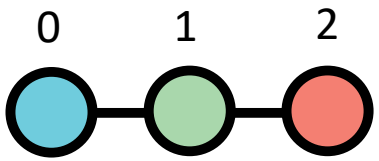
Change due to 2 reactions

Change due to 3 reactions

...

Theory I – linear CRNs

$$[\mathrm{X_k}] = [\mathrm{X_k}]_0 + t\,(\mathbb{J}\,[\mathbf{X}]_0) + t^2\frac{(\mathbb{J}^2[\mathbf{X}]_0)}{2} + t^3\frac{(\mathbb{J}^3[\mathbf{X}]_0)}{3!} + \ldots$$



Change due to 1 reaction

Change due to 2 reactions

Change due to 3 reactions

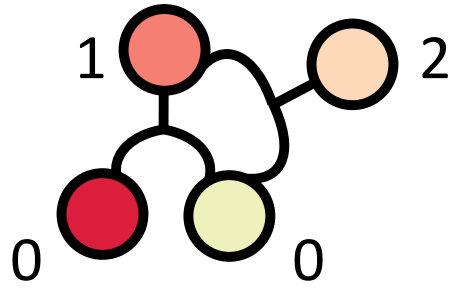
...

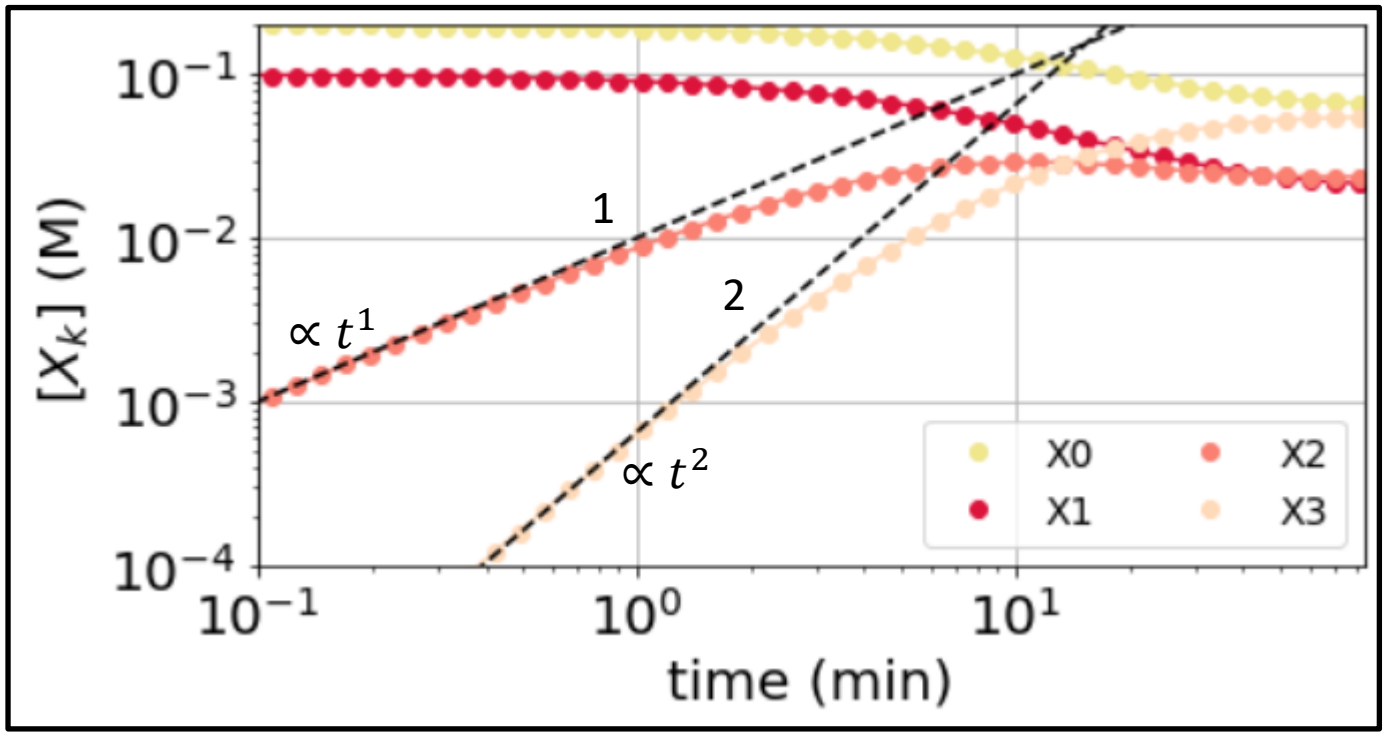
$$\mathbb{J}^2 = \begin{pmatrix} k_1^+(k_1^+ + k_1^-) & -k_1^-(k_1^+ + k_1^- + k_2^+) & \boxed{k_1^-\,k_2^-} \\ -k_1^+(k_1^+ + k_1^- + k_2^+) & k_1^+k_1^- + (k_1^- + k_2^+)^2 + k_2^+k_2^- & -k_2^-(k_1^- + k_2^+ + k_2^-) \\ \boxed{k_1^+k_2^+} & -k_2^+(k_1^+ + k_2^+ + k_2^-) & k_2^-(k_2^+ + k_2^-) \end{pmatrix}$$

$$\mathbb{J} = \begin{pmatrix} -k_1^+ & k_1^- & \boxed{0} \\ k_1^+ & -k_1^- - k_2^+ & k_2^- \\ \boxed{0} & k_2^+ & k_2^- \end{pmatrix}$$

Reaction steps for nonlinear networks

$$X_0 + X_1 \rightleftarrows X_2 \qquad X_2 + X_0 \rightleftarrows X_3$$





Kinetic exponents: General CRNs

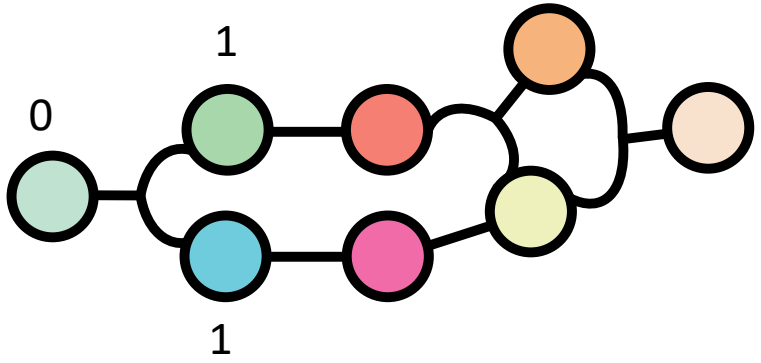
Nonlinear dynamics still has series expansion
+ pathlike interpretation

$$d_t[\mathbf{X}] = \hat{\mathbb{J}}([\mathbf{X}])[\mathbf{X}]$$

$$[\mathrm{X_k}] = a_{\kappa_k}\, t^{\kappa_k} + a_{\kappa_k+1}\, t^{\kappa_k+1} + \ \ldots$$

Kinetic Exponent law

Kin. Exp. product =
1+ $\sum$ Kin. Exp. reactants



Kinetic exponents: General CRNs

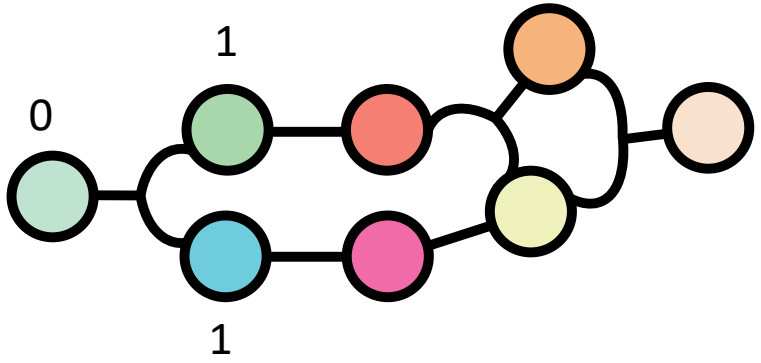
Nonlinear dynamics still has series expansion + pathlike interpretation

$$d_t[\mathbf{X}] = \hat{\mathbb{J}}([\mathbf{X}])[\mathbf{X}]$$

$$[\mathrm{X_k}] = a_{\kappa_k}\, t^{\kappa_k} + a_{\kappa_k+1}\, t^{\kappa_k+1} + \ \ldots$$

Kinetic Exponent law

$\kappa(\text{product}) =$
$1 + \sum \nu_{(\text{reactant})}\, \kappa(\text{reactant})$



Kinetic exponents: General CRNs

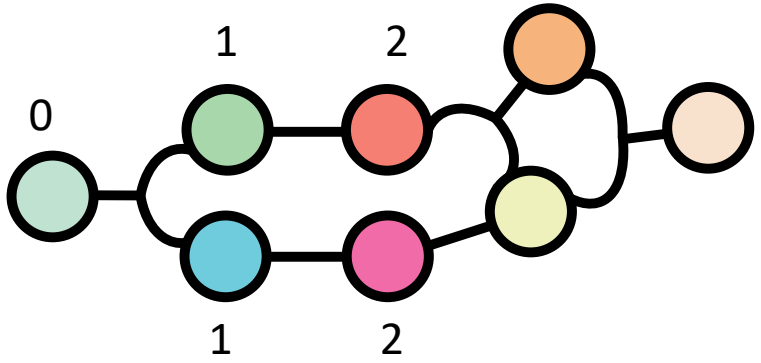
Nonlinear dynamics still has series expansion + pathlike interpretation

$$d_t[\mathbf{X}] = \hat{\mathbb{J}}([\mathbf{X}])[\mathbf{X}]$$

$$[\mathrm{X_k}] = a_{\kappa_k}\, t^{\kappa_k} + a_{\kappa_k+1}\, t^{\kappa_k+1} + \ \ldots$$

Kinetic Exponent law

$\kappa(\text{product}) =$
$1 + \sum \nu_{(\text{reactant})}\, \kappa(\text{reactant})$



Kinetic exponents: General CRNs

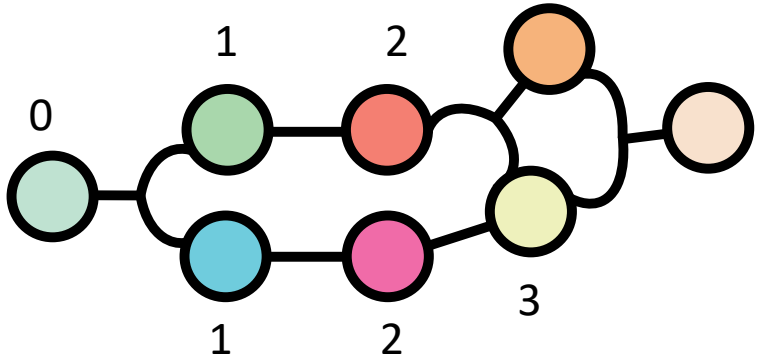
Nonlinear dynamics still has series expansion
+ pathlike interpretation

$$d_t[\mathbf{X}] = \hat{\mathbb{J}}([\mathbf{X}])[\mathbf{X}]$$

$$[\mathrm{X_k}] = a_{\kappa_k}\, t^{\kappa_k} + a_{\kappa_k+1}\, t^{\kappa_k+1} + \ \ldots$$

Kinetic Exponent law

$\kappa(\text{product}) =$
$1 + \sum \nu_{(\text{reactant})}\, \kappa(\text{reactant})$



Kinetic exponents: General CRNs

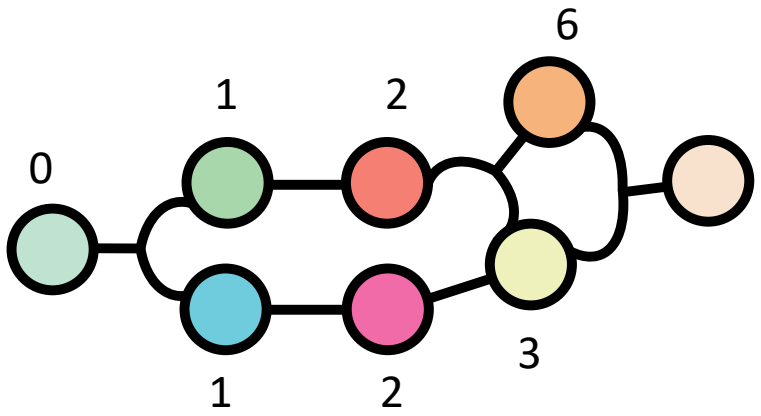
Nonlinear dynamics still has series expansion
+ pathlike interpretation

$$d_t[\mathbf{X}] = \hat{\mathbb{J}}([\mathbf{X}])[\mathbf{X}]$$

$$[\mathrm{X_k}] = a_{\kappa_k}\, t^{\kappa_k} + a_{\kappa_k+1}\, t^{\kappa_k+1} + \ \ldots$$

Kinetic Exponent law

$\kappa(\text{product}) =$
$1 + \sum \nu_{(\text{reactant})}\, \kappa(\text{reactant})$



Kinetic exponents: General CRNs

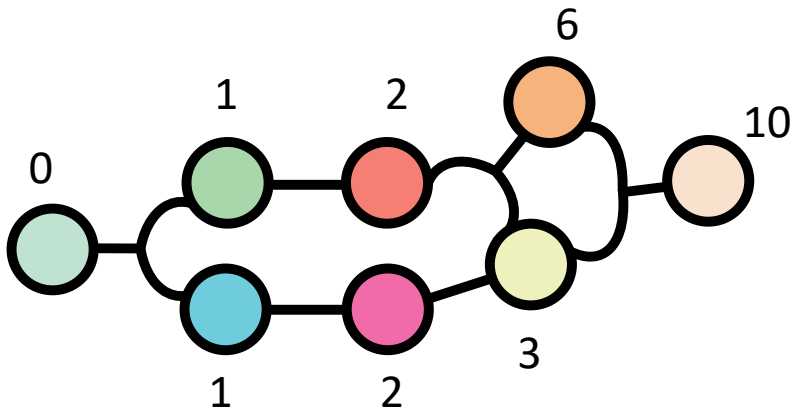
Nonlinear dynamics still has series expansion
+ pathlike interpretation

$$d_t[\mathbf{X}] = \hat{\mathbb{J}}([\mathbf{X}])[\mathbf{X}]$$

$$[\mathrm{X_k}] = a_{\kappa_k}\, t^{\kappa_k} + a_{\kappa_k+1}\, t^{\kappa_k+1} + \ \ldots$$

Kinetic Exponent law

$\kappa(\text{product}) =$
$1 + \sum \nu_{(\text{reactant})}\, \kappa(\text{reactant})$



Kinetic exponents: General CRNs

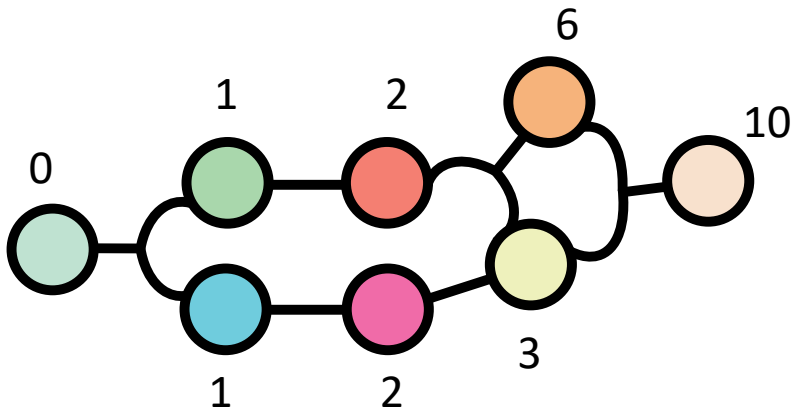
Nonlinear dynamics still has series expansion
+ pathlike interpretation

$$d_t[\mathbf{X}] = \hat{\mathbb{J}}([\mathbf{X}])[\mathbf{X}]$$

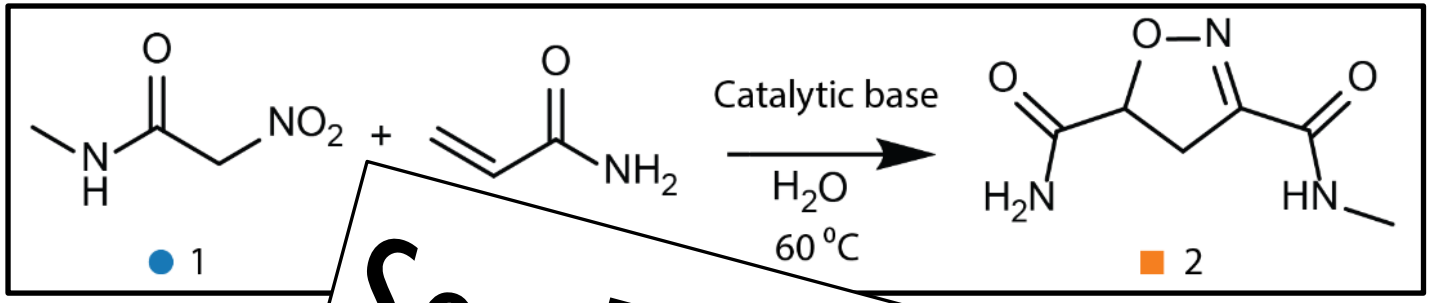
$$[\mathrm{X_k}] = a_{\kappa_k}\, t^{\kappa_k} + a_{\kappa_k+1}\, t^{\kappa_k+1} + \ \ldots$$

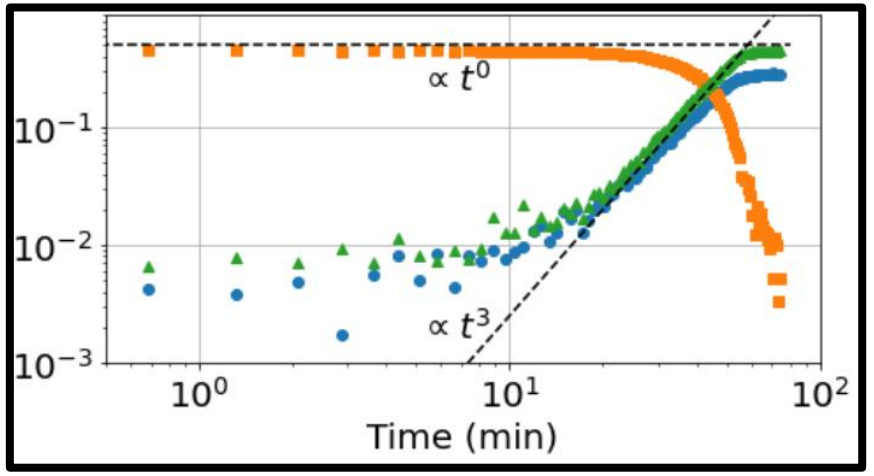
Kinetic Exponent law

$\kappa(\text{product}) =$
$1 + \sum \nu_{(\text{reactant})}\, \kappa(\text{reactant})$

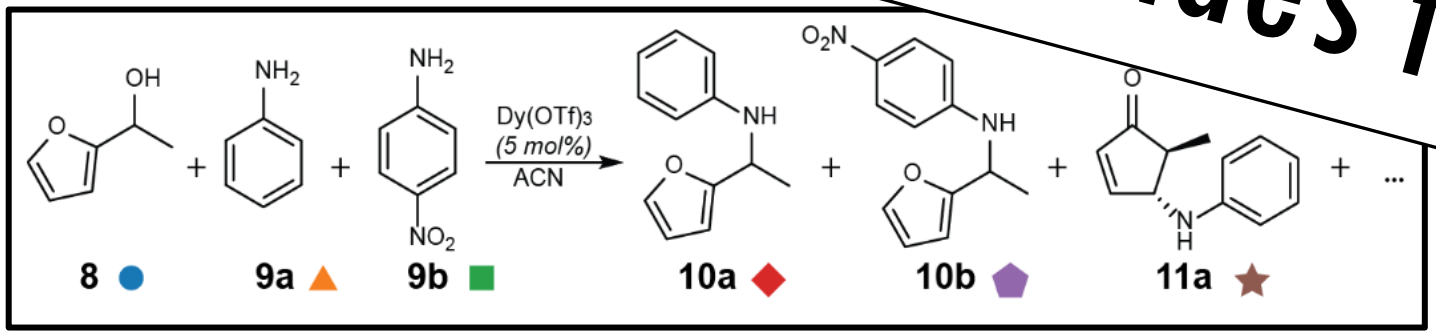


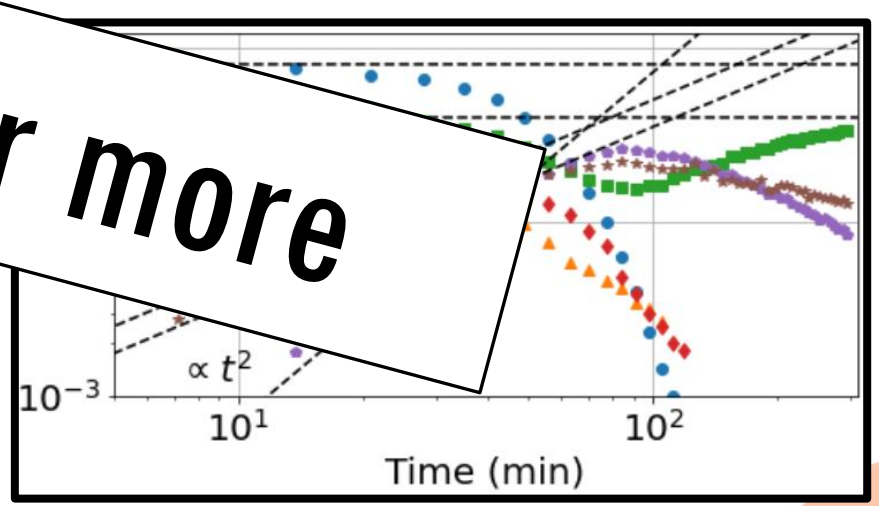
Kinetic exponents are extractable from real data



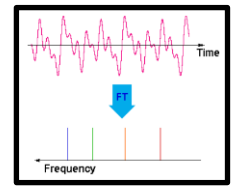


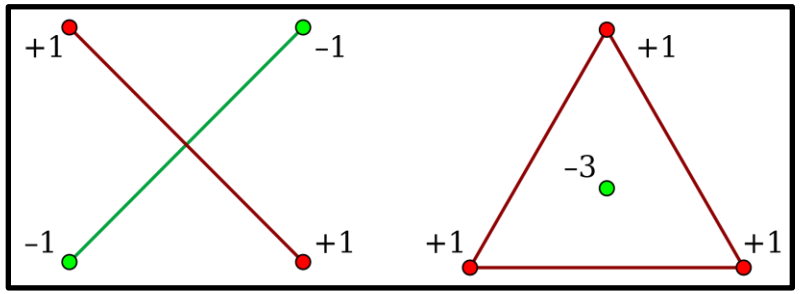
See Doubice slides for more



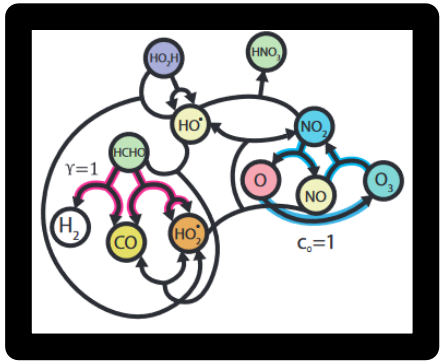


Some measurable CRN properties

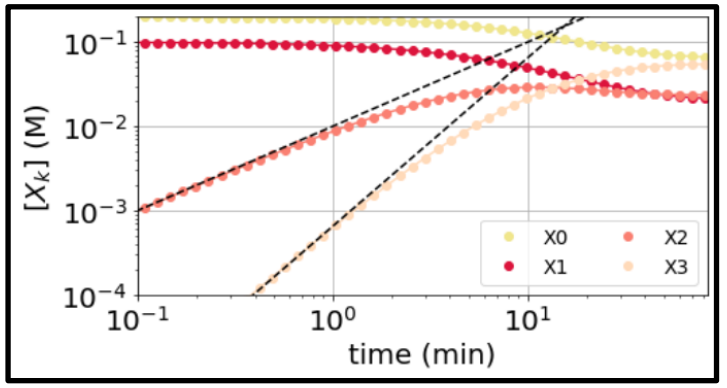




Radon partitions



Data dimension & conservation laws



Kinetic exponents

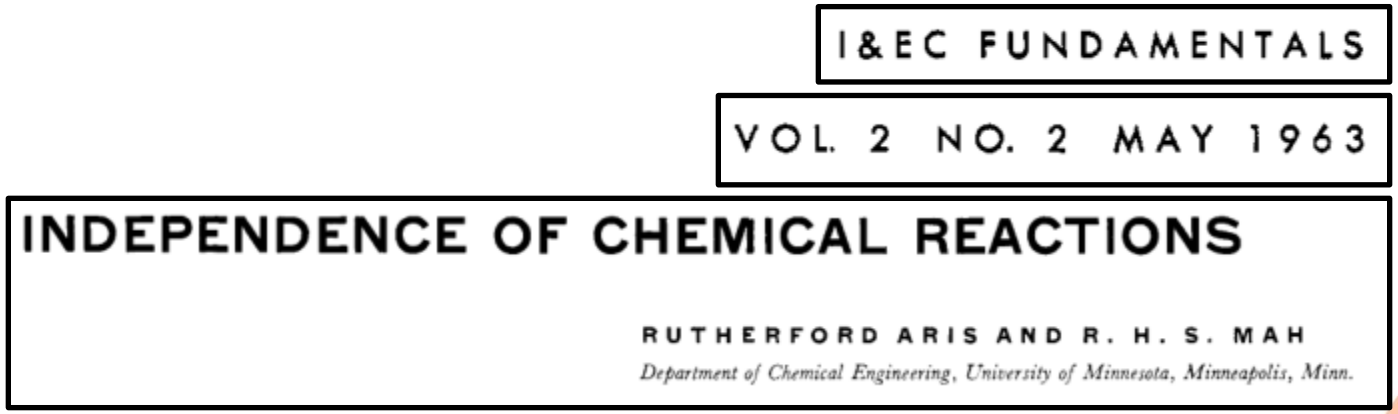
Data dimension – some history

Basic rough idea (d):

d = # independent reactions = # independent species

d = 'effective' rank* of (mean-subtracted) data

*In the linear algebra sense



Data dimension

Basic rough idea (d):

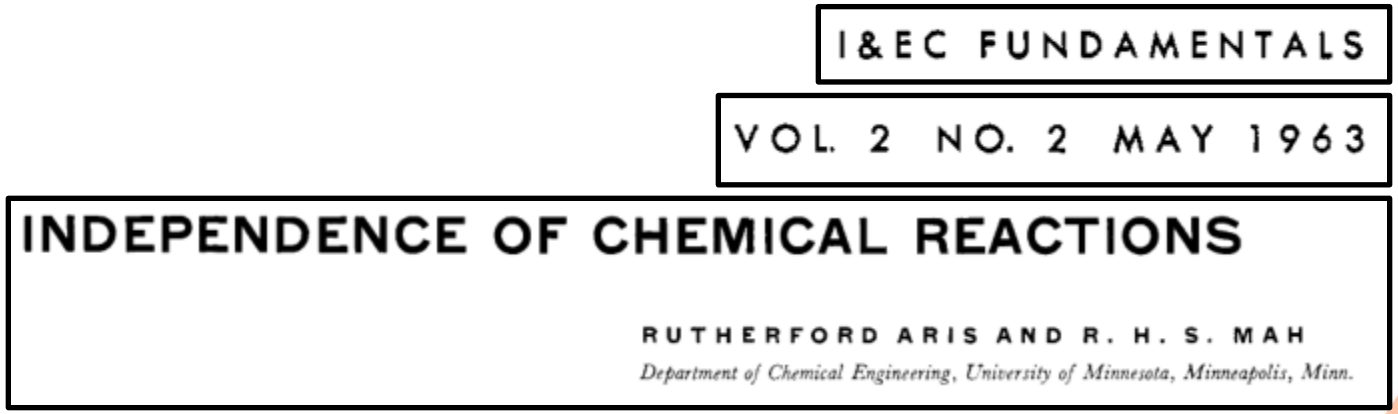
d = # independent reactions = # independent species

d = 'effective' rank of (mean-subtracted) data

$$d = s - l = r - c$$

#species - # conservation laws = # reactions - # cycles

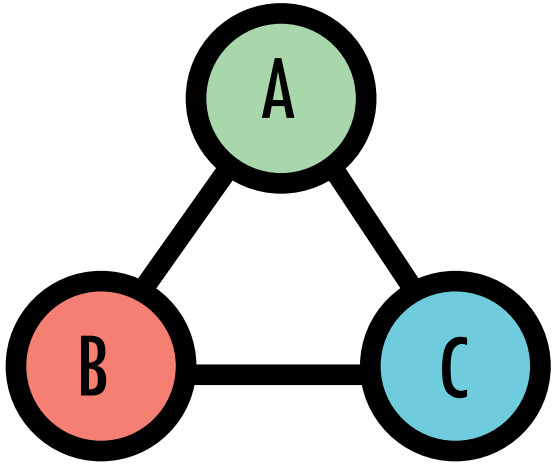
Found very limited experimental adoption.



The first structural law

$d = s - l = r - c$

#species - # conservation laws = # reactions - # cycles



$A \rightleftarrows B \rightleftarrows C \rightleftarrows A$

$s = 3$ (A,B,C)

$\ell = 1$ ([A] + [B] + [C] = Constant)

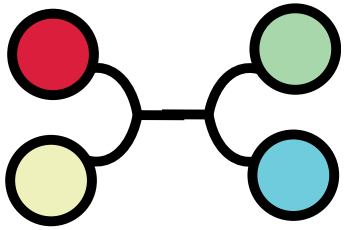
$r = 3$

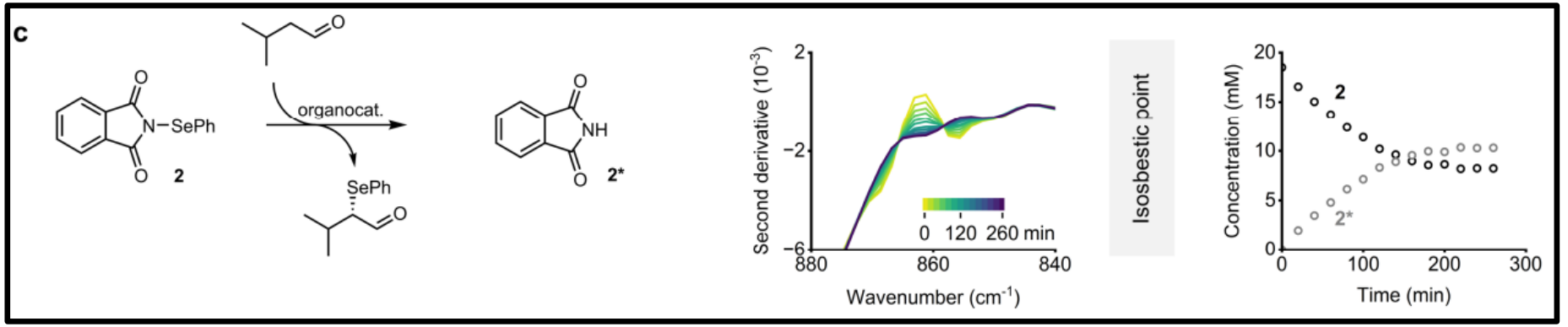
$c = 1$

Used in a limited sense: isosbestic points (d=1)

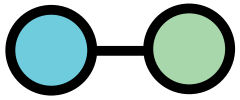
- Spectral overlap (reactant(s), product(s))
- **1d** transformation (chemical, physical)

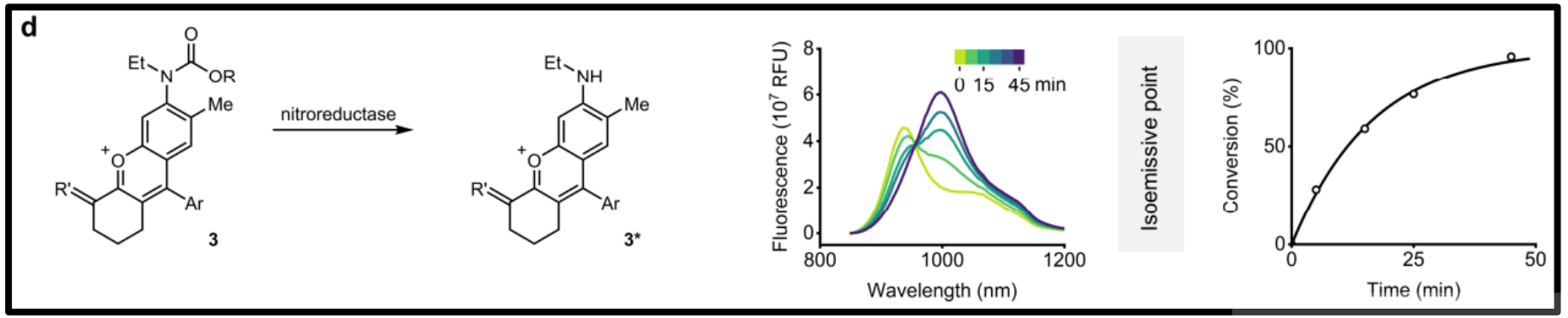
d=1 indicates particularly simple transformation, no side reactions*





isosbestic

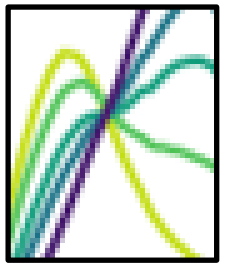


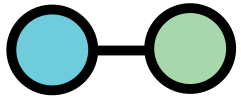


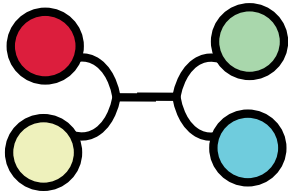
isoemmissive

Data dimension

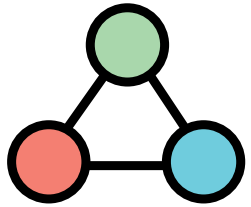
d=1
(isosbestic point)

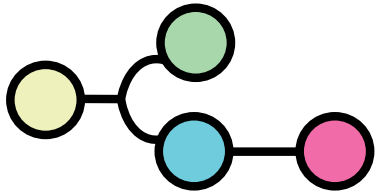






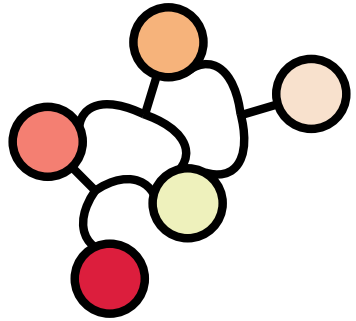
d=2
(Isosbestic ~~line~~ curve)





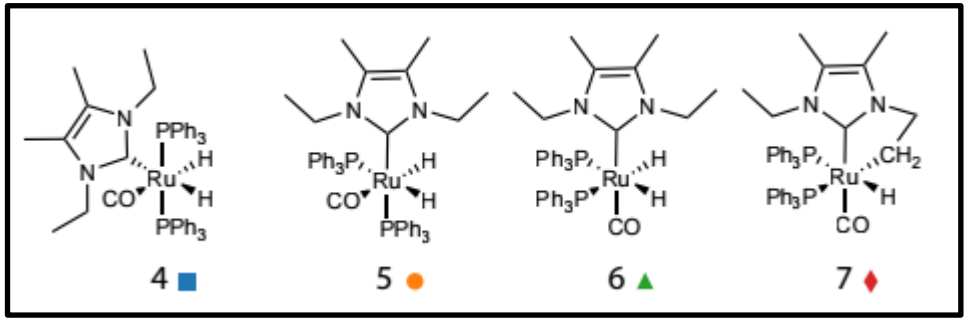
d=3

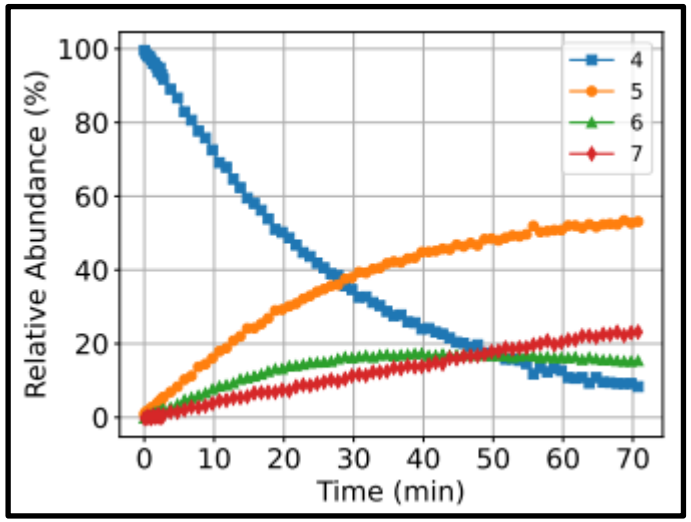
Increasingly hard **to see** directly. (low) d can in general be estimated via Singular Value Decomposition / PCA / ...



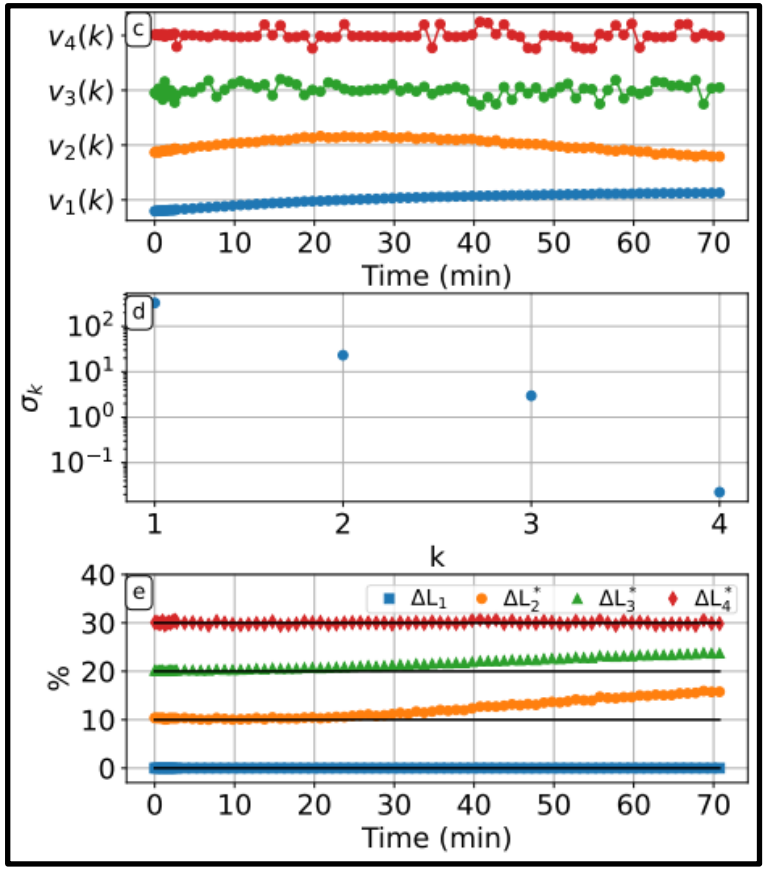
d=4

Quick example



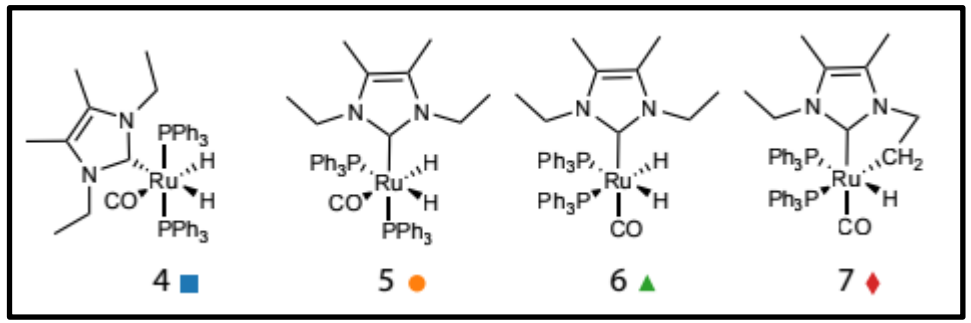


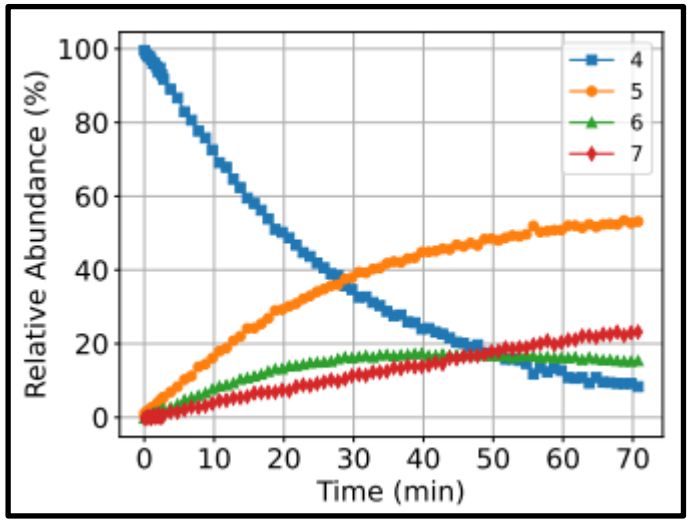
SVD →



full data d = 2 , on short timescales d=1

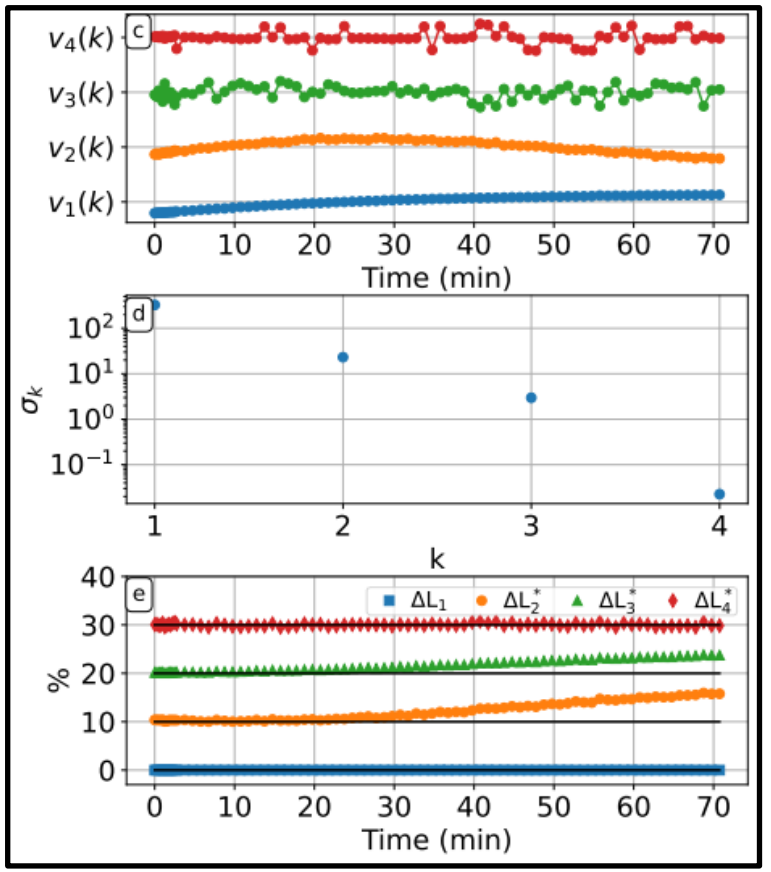
Quick example





full data d = 2 , on short timescales d=1

SVD →



Problem: classical theory naively predicts d=3 and ignores time-dependence / resolution d !

Data dimension vs the real world

Basic rough idea (d):

d = # independent reactions = # independent species = 'rank' of (mean-subtracted) data

Some problems with the naive theory:

- Techniques often don't see all species
- Discernable dimension ("rank") of data depends on resolution (time, concentration, # variables, . . .)
- Chemical phenomenology (e.g. phase transitions, collinear reactions) can alter d
- dim(data) ≠ dim(CRN) ?

↓

Emergent conservation laws, hidden currents, new theory needed

[Submitted on 15 Jun 2023 (v1), last revised 8 Apr 2024 (this version, v3)]

On data and dimension in chemistry I -- irreversibility, concealment and emergent conservation laws

Alex Blokhuis, Martijn van Kuppeveld, Daan van de Weem, Robert Pollice

Research Article | Open Access

Case Studies of Dimensionality in Chemical Data

Dr. Alex Blokhuis, Dr. Robert Pollice

First published: 24 December 2024 | https://doi.org/10.1002/ejoc.202400949

Theory to bridge experimental resolution and chemical phenomenology

[Submitted on 15 Jun 2023 (v1), last revised 8 Apr 2024 (this version, v3)]

On data and dimension in chemistry I -- irreversibility, concealment and emergent conservation laws

Alex Blokhuis, Martijn van Kuppeveld, Daan van de Weem, Robert Pollice

arXiv:2306.09553

***In review,* PRX**

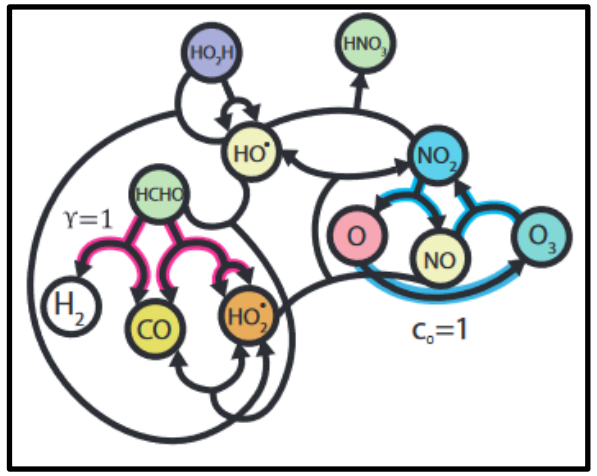
Research Article | Open Access

Case Studies of Dimensionality in Chemical Data

Dr. Alex Blokhuis, Dr. Robert Pollice

First published: 24 December 2024 | https://doi.org/10.1002/ejoc.202400949

EurJOC
European Journal of Organic Chemistry

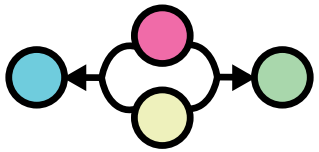


Lot more indices + laws

Example: Co-production conservation laws

Emergent conservation laws, hidden currents, new theory needed

s – l = r – c ?

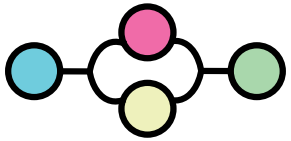


d

$C \leftarrow A + B \longrightarrow D$

$l_s = 2, l_{irr} = 1$

1



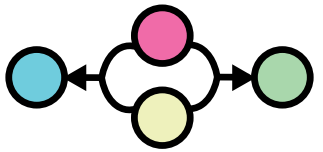
$C \rightleftarrows A + B \rightleftarrows D$

$l_s = 2, \mathrm{d} = 2$

2

Co-production conservation laws

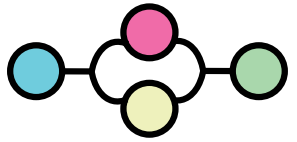
$$L = \kappa_2^+[C] - \kappa_1^+[D]$$



$C \leftarrow A + B \rightarrow D$
$l_s = 2, l_{irr} = 1$

d

1



$C \rightleftarrows A + B \rightleftarrows D$
$l_s = 2, \mathrm{d} = 2$

2

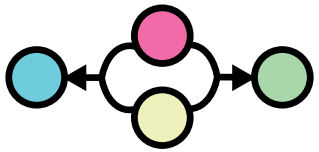
Co-production conservation laws

Merge collinear reactions, now reactions are genuinely independent (vis-a-vis dynamics).

$$A + B \longrightarrow pC + (1-p)D$$

$$\Upsilon = \mathfrak{d}_{\bullet} + \Lambda_{\bullet}, \tag{47}$$

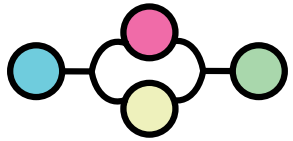
where
Υ: co-production index, # (collinear) co-production relations
$\mathfrak{d}_{\bullet}$: # co-production emanants (emergent conservation laws),
$\Lambda_{\bullet}$: # broken cycles.



$C \leftarrow A + B \longrightarrow D$
$l_s = 2, l_{irr} = 1$

d

1



$C \rightleftarrows A + B \rightleftarrows D$
$l_s = 2, \mathrm{d} = 2$

2

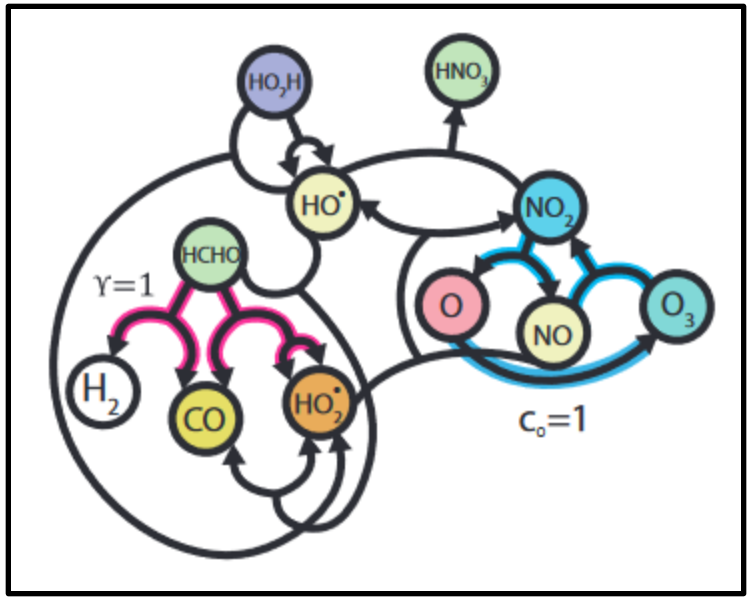
Co-production conservation laws

Merge collinear reactions, now reactions are genuinely independent (vis-a-vis dynamics).

$$A + B \longrightarrow pC + (1-p)D$$

$$\Upsilon = \mathfrak{d}_{\bullet} + \Lambda_{\bullet}, \qquad (47)$$

where
Υ: co-production index, # (collinear) co-production relations
$\mathfrak{d}_{\bullet}$: # co-production emanants (emergent conservation laws),
$\Lambda_{\bullet}$: # broken cycles.



Dimension and phase behavior

$$X(s) \rightleftarrows X_1(aq) \qquad [X_1]_{eq} = K_s \quad \text{solubility}$$

$$X_1 + X_1 \rightleftarrows X_2 \qquad [X_2]_{eq} = K_2 [X_1]_{eq}^2$$

$$\vdots \qquad\qquad \vdots$$

$$X_1 + X_{n-1} \rightleftarrows X_n \qquad [X_n]_{eq} = K_n [X_1]_{eq}^n$$

$$\emptyset \rightarrow X_1$$

$$s = n, \ell = 0$$
$$d = s$$

Suppose we slowly add X_1

s −dimensional data

Dimension and phase behavior

$$X(s) \rightleftarrows X_1(aq)$$
$$X_1 + X_1 \rightleftarrows X_2$$
$$\vdots$$
$$X_1 + X_n \rightleftarrows X_{n+1}$$

$$\emptyset \rightarrow X_1$$

$$[X_1]_{eq} = K_s \quad \text{solubility}$$
$$[X_2]_{eq} = K_2 [X_1]_{eq}^2 = K_2 K_s^2$$
$$\vdots$$
$$[X_n]_{eq} = K_n [X_1]_{eq}^n = K_n K_s^n$$

$$s = n, \ell = 0$$
$$d = s$$

Suppose we slowly add X_1
In the presence of a phase equilibrium,
0-dimensional data!

Phase behavior can dramatically alter data dimension
Phases can be small (e.g. micelles), phase behavior often goes unnoticed!

Dimension and phase behavior

$$X(s) \rightleftarrows X_1(aq)$$

$$X_1 + X_1 \rightleftarrows X_2$$

$$\vdots$$

$$X_1 + X_n \rightleftarrows X_{n+1}$$

$$\emptyset \rightarrow X_1$$

$$[X_1]_{eq} = K_s \quad \text{solubility}$$

$$[X_2]_{eq} = K_2 [X_1]_{eq}^2 = K_2 K_s^2$$

$$\vdots$$

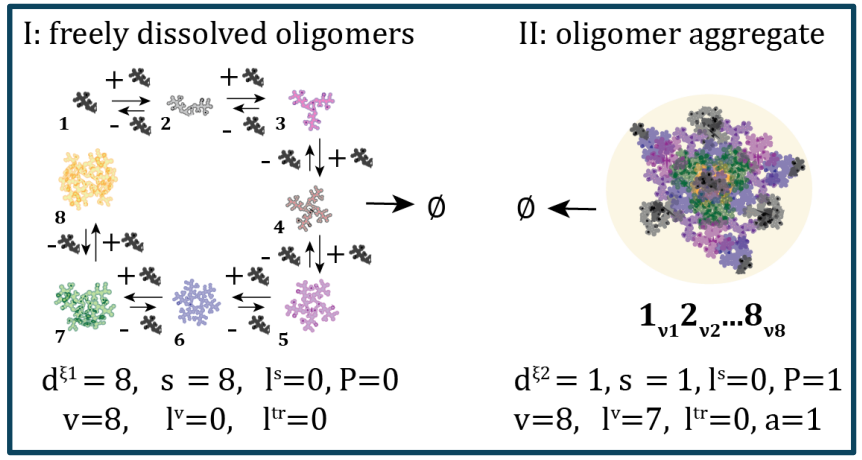
$$[X_n]_{eq} = K_n [X_1]_{eq}^n = K_n K_s^n$$

$$s = n, \ell = 0$$
$$d = s$$

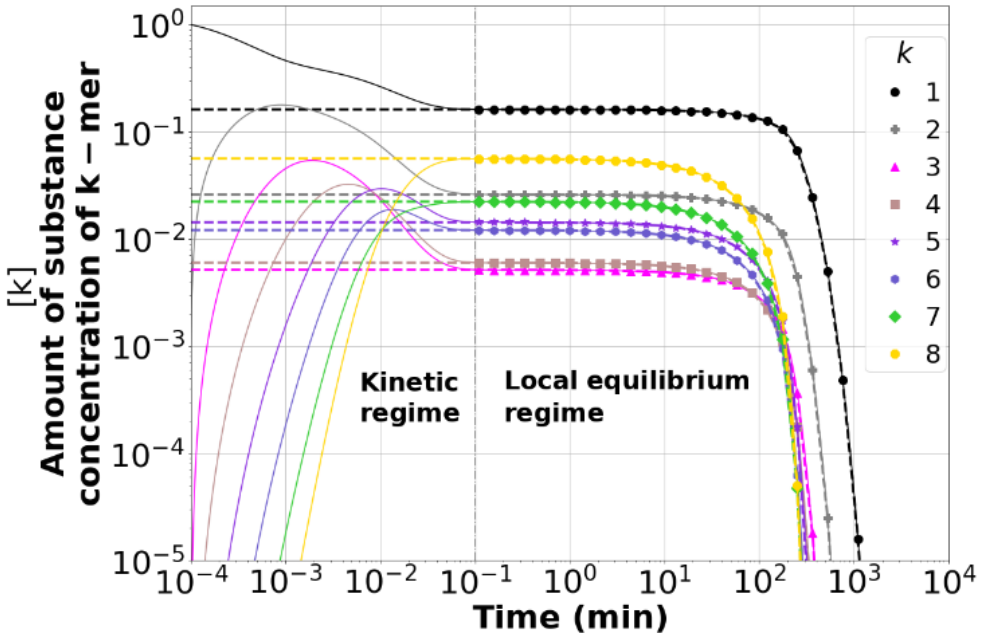
Suppose we slowly add X_1
In the presence of a phase equilibrium,
0-dimensional data!

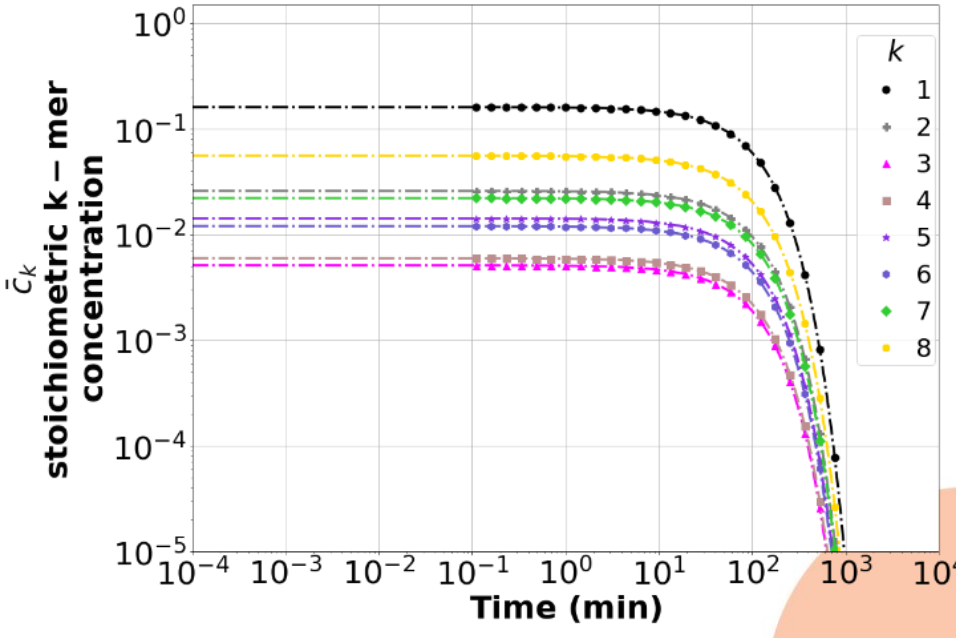
Phase behavior can dramatically alter data dimension
Phases can be small (e.g. micelles), phase behavior often goes unnoticed!

Very different behavior, same observables



In dynamic combinatorial chemistry, one **oftentimes** observes very low-dimensional data (d=1,2) in spite of there being many species that can be isolated (e.g. by HPLC-MS)





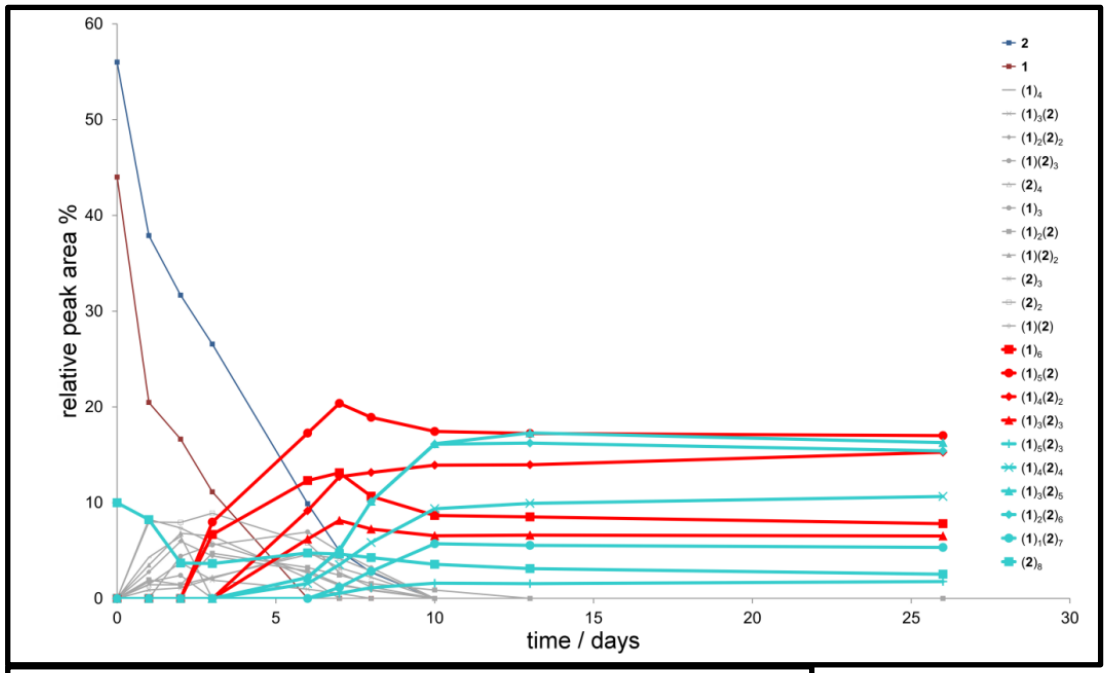
Example from systems chemistry

Typical data: UPLC chromatogram with distribution of oligomers 1 – 18

Single building block: $l_s = 2$, two building blocks: $l_s = 3$,

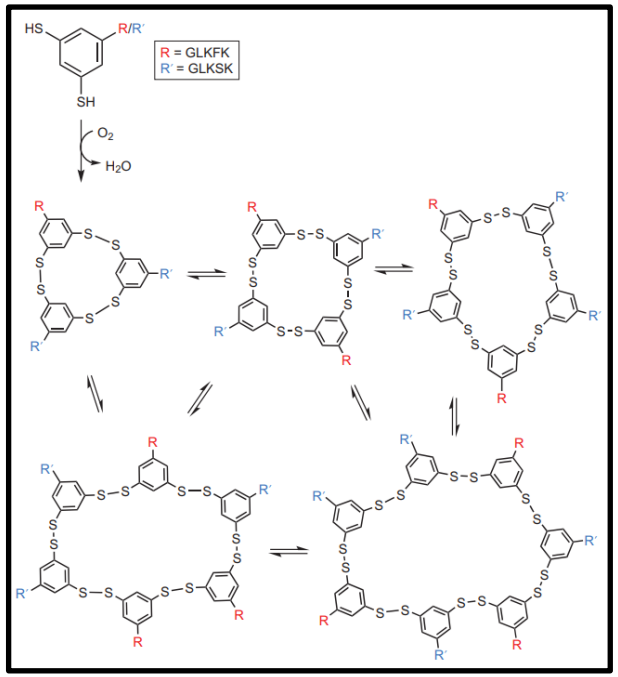
In practice, we systematically find low-dimensional data (d=2, 3 , $l > 10$)

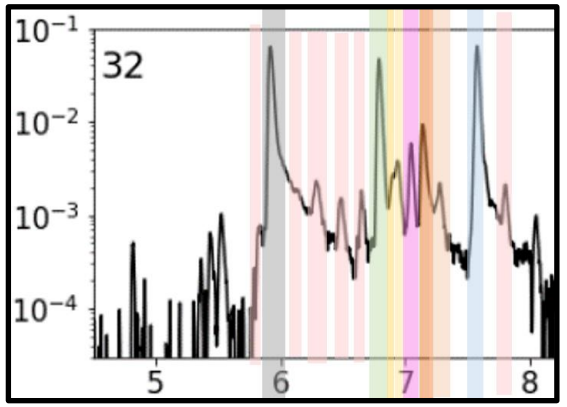
How can that be?

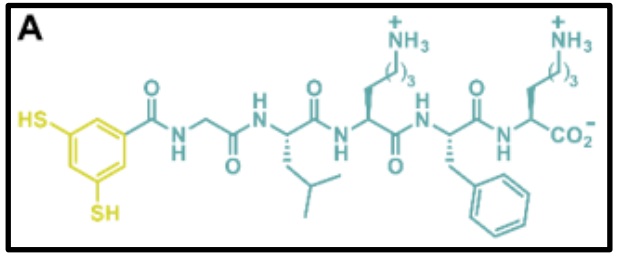


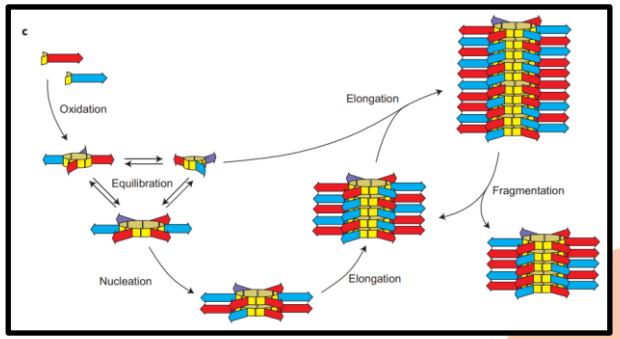
Diversification of self-replicating molecules

Jan W. Sadownik, Elio Mattia, Piotr Nowak and Sijbren Otto*









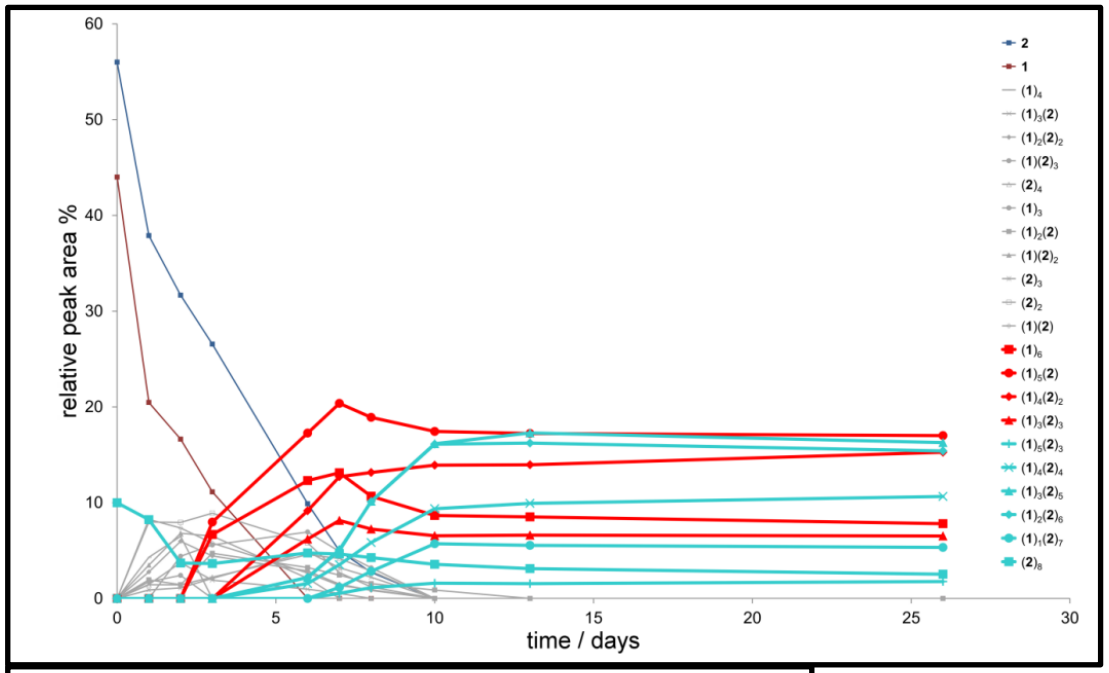
Example from systems chemistry

Typical data: UPLC chromatogram with distribution of oligomers 1 – 18

Single building block: $l_s = 2$, two building blocks: $l_s = 3$,

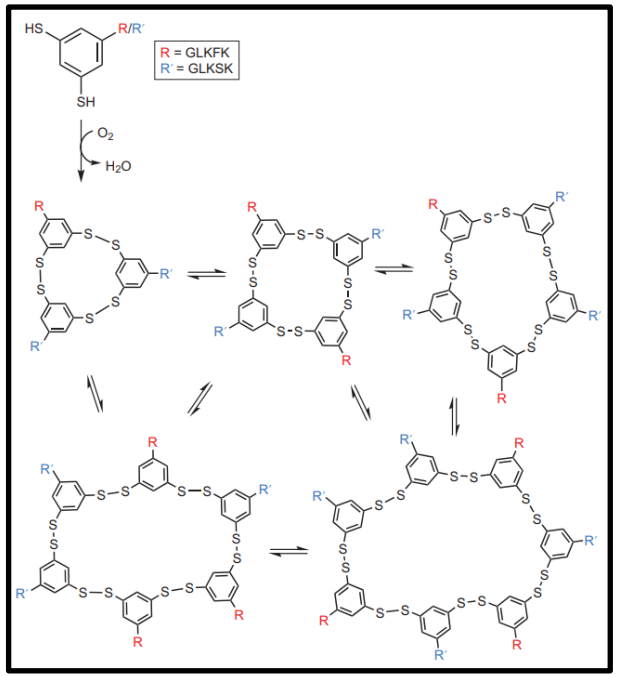
In practice, we systematically find low-dimensional data (d=2, 3 , $l > 10$)

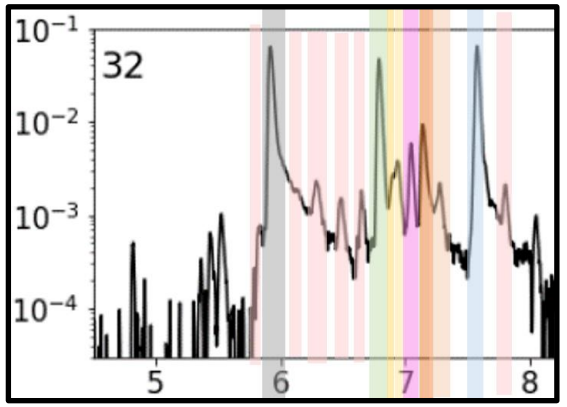
How can that be?

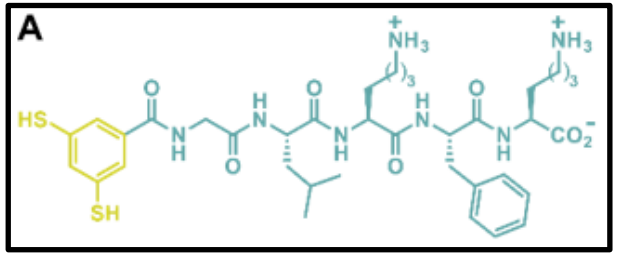


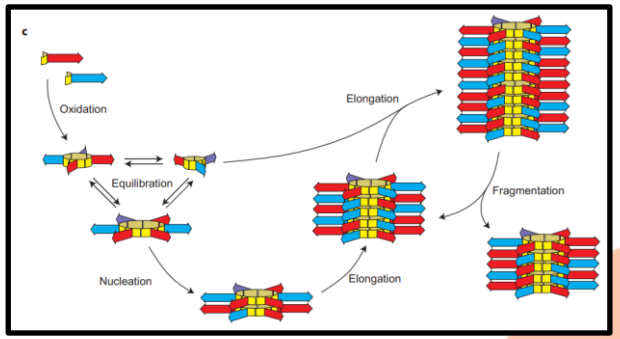
Diversification of self-replicating molecules

Jan W. Sadownik, Elio Mattia, Piotr Nowak and Sijbren Otto*









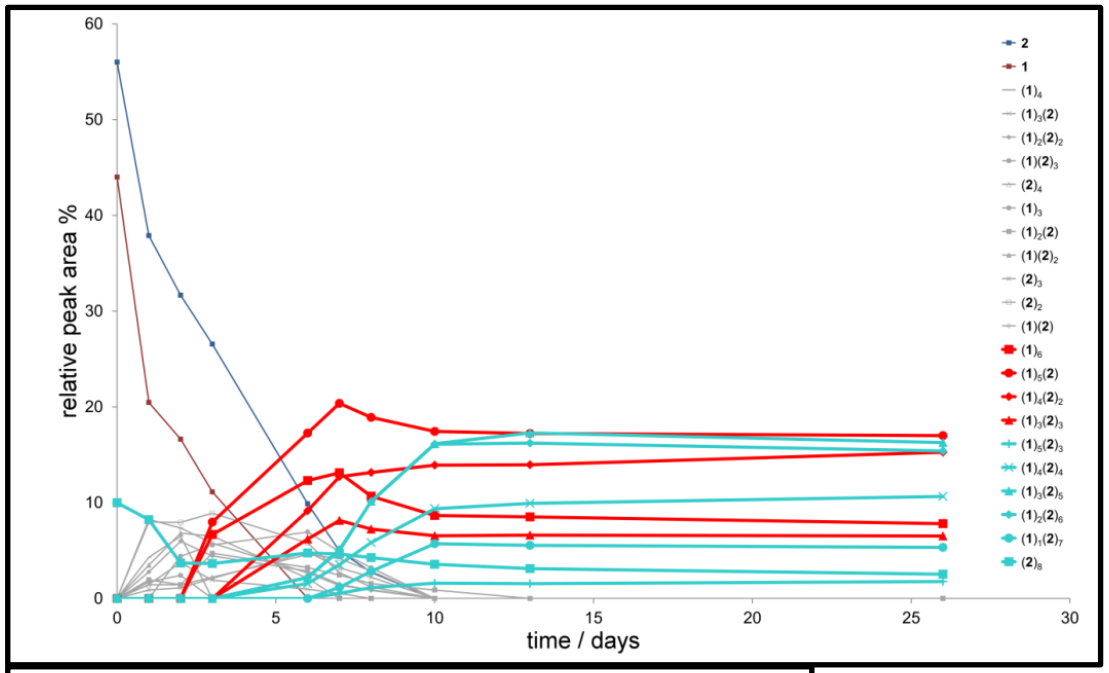
Example from systems chemistry

Typical data: UPLC chromatogram with distribution of oligomers $1-18$

Single building block: $l_s = 2$, two building blocks: $l_s = 3$,

In practice, we systematically find low-dimensional data (d=2, 3 , $l > 10$)

How can that be? ↦ Hidden phase behavior

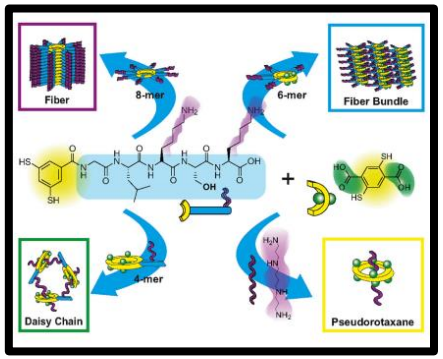


Diversification of self-replicating molecules

Jan W. Sadownik, Elio Mattia, Piotr Nowak and Sijbren Otto*

Caught in the Act: Mechanistic Insight into Supramolecular Polymerization-Driven Self-Replication from Real-Time Visualization

Sourav Maity, Jim Ottelé, Guillermo Monreal Santiago, Pim W. J. M. Frederix, Peter Kroon, Omer Markovitch, Marc C. A. Stuart, Siewert J. Marrink,* Sijbren Otto,* and Wouter H. Roos*



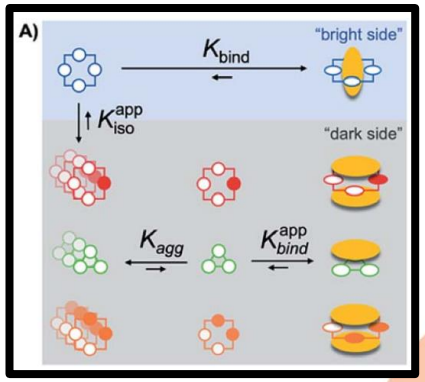
Article

Stoichiometry alone can steer supramolecular systems on complex free energy surfaces with high selectivity

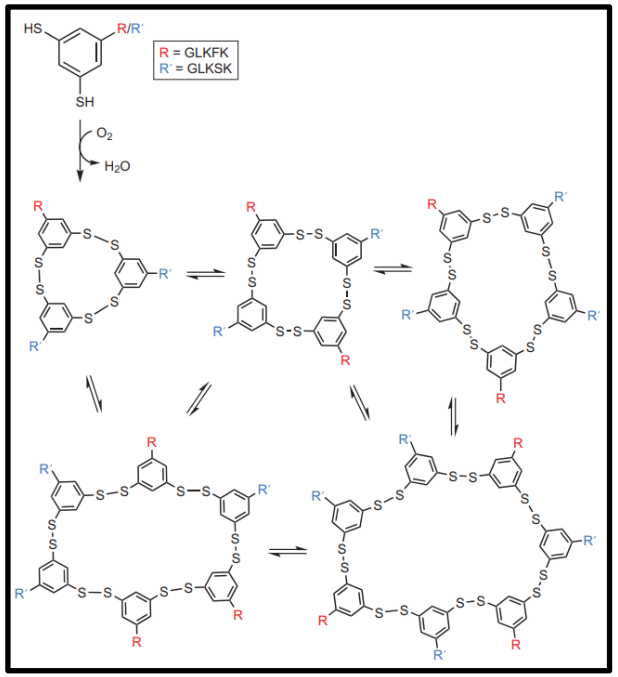
Dávid Komáromy,[1,*] Theodora Tiemersma-Wegman,[1] Johan Kemmink,[1] Giuseppe Portale,[2] Paul R. Adamski,[1] Alex Blokhuis,[1] Friso S. Aalbers,[3] Ivana Marić,[1] Guillermo Monreal Santiago,[1] Jim Ottelé,[1] Ankush Sood,[1] Vittorio Saggiomo,[1,4] Bin Liu,[1] Pieter van der Meulen,[1] and Sijbren Otto[1,5,*]

The dark side of disulfide-based dynamic combinatorial chemistry†

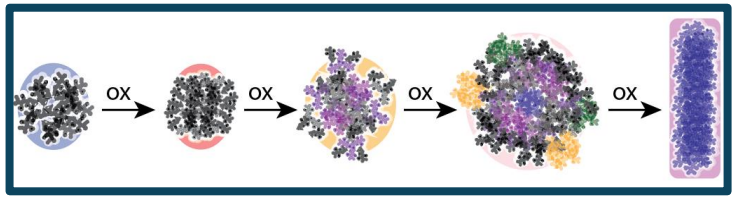
Mélissa Dumartin, [a] Jean Septavaux, [ab] Marion Donnier-Maréchal,[a] Emeric Jeamet,[a] Elise Dumont, [cd] Florent Perret, [*a] Laurent Vial [*a] and Julien Leclaire [*a]

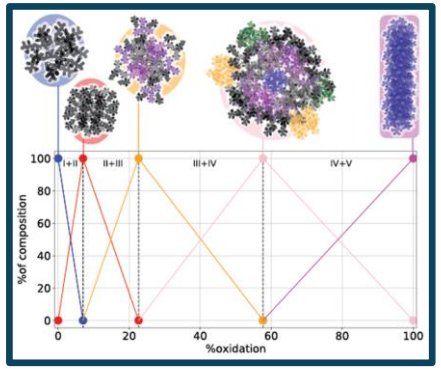


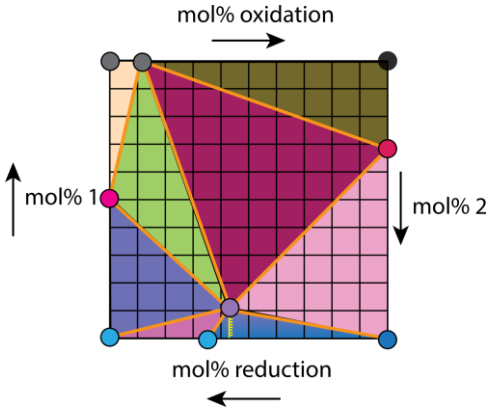
Emergent simplicity in chemistry due to (psuedo)phases

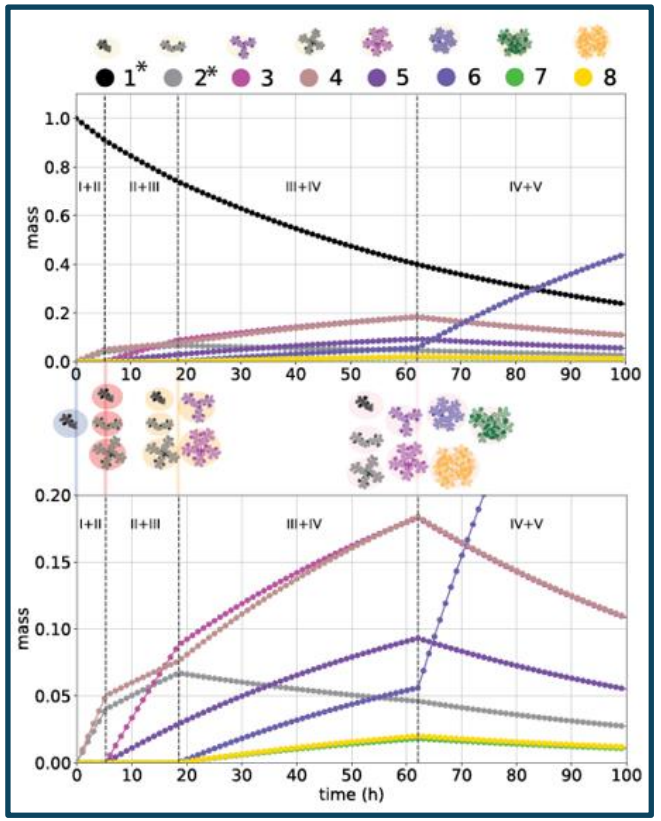


Large molecular CRNS
Many ODEs









Phase rules, phase diagrams

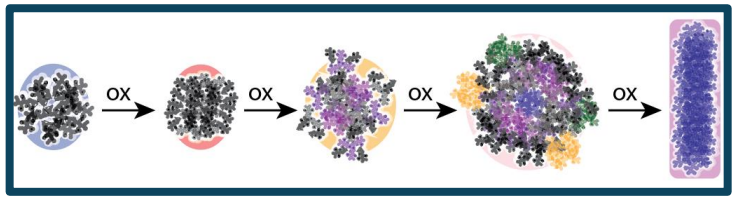
Emergent simplicity in chemistry due to (psuedo)phases

A.B., Y. Geiger, S. Otto. When aggregation becomes the norm. *invited, in preparation* , J. Am. Chem. Soc.

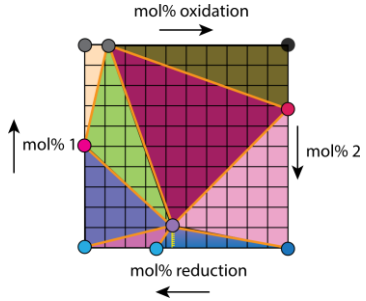
Many chemical systems exhibit signatures and prerequisites of hidden phase behavior, e.g.

- Common reactions in organic chemistry
- Concentrated salt solutions
- Formose reaction
- Oligopeptide solutions

⋮



This profoundly alters their description, and how we can optimize them for a given task.



Large molecular CRNS
Many ODEs

→

Phase rules, phase diagrams

Thank you

Ottolab

Nicola Vassena

Martijn van Kuppeveld

Hermanslab

Yannick Geiger

Robert Pollice

Daan van de Weem

Oriane Cosker

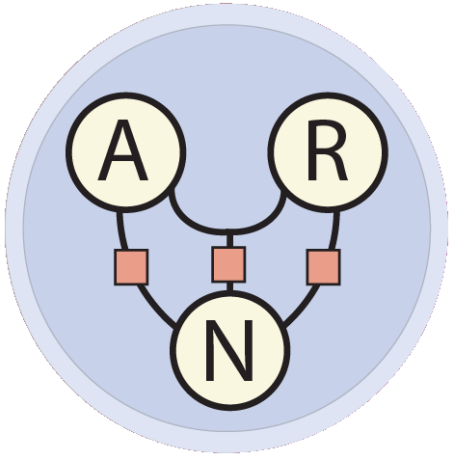
Promo

Join our interdisciplinary autocatalysis seminar

researchseminars.org/seminar/AutocatalysisRN

@ARNseminar

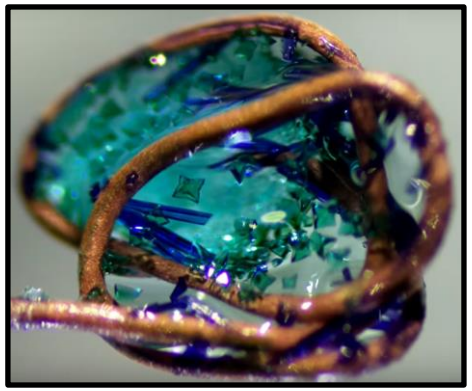
w. Praful Gagrani, Nicola Vassena, Wei-Hsiang Lin



@alexblokhuis

@AlexBlokhuis

@alexblokhuis



Daily video microscopy chemistry experiments