MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **95** (2026) 347–404
ISSN: 0340–6253
doi: 10.46793/match.95-2.15425

Atom Transition Networks and Isotope Labeling Patterns in Large Chemical Reaction Networks

Richard Golnik[1,*], **Peter F. Stadler**[1–7], **Thomas Gatter**[1]

[1] *Bioinformatics Group, Department of Computer Science, Leipzig University, Härtelstraße 16–18, D-04107, Leipzig, Germany*

[2] *Interdisciplinary Center for Bioinformatics & Zuse School for Embedded and Composite Artificial Intelligence (SECAI) & Center for Scalabale Data Analytics and Artificial Intelligence, Leipzig University, D-04107, Leipzig, Germany*

[3] *Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, D-04103, Leipzig, Germany*

[4] *Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090, Vienna, Austria*

[5] *Facultad de Ciencias, Universidad National de Colombia, Bogotá, Colombia*

[6] *Center for non-coding RNA in Technology and Health, University of Copenhagen, Ridebanevej 9, DK-1870, Frederiksberg, Denmark*

[7] *Santa Fe Institute, 1399 Hyde Park Rd., Santa Fem, NM, 87501, USA*

{richard, studla, thomas}@bioinf.uni-leipzig.de

(Received June 23, 2025)

*Corresponding author.

Abstract

Atom transition networks (ATNs) provide the fine-grained, atom-level description of a chemical reaction network that is required for a detailed, mechanistic understanding of multi-step reactions and for the practical analysis of isotope labeling experiments. Conceptually, ATNs are determined completely by (i) the reaction-level description of a chemical reaction networks, (ii) the atom-to-atom map for each constituent reaction, and (iii) fluxes specifying the relative contribution of different reactions to the turnover of individual reactants. The construction of ATNs, and thus the analysis of isotope tracing experiments, is a notoriously difficult and time-consuming task that is aggravated by symmetries in molecules and reactions. Starting from the atom-to-atom map of a reaction we first derive a transition matrix that exactly describes the propagation of the label from reactants to products, taking into account the relevant symmetries. These are subsequently combined into a ATN. Assuming a steady-state flux F through the chemical reaction network we derive a system of affine differential equations describing how the reaction network is flooded by labeled atoms from an external reservoir. This leads to a unique asymptotically stable steady-state distribution of labels that can be computed by solving a non-singular system of linear equations. Linear combinations of these single-atom labeling patterns also solve the problem of computing the enrichment of multiple, simultaneous labels, albeit without providing information on correlations between distinct labels. In particular, we present a simple and complete solution for an important special case of isotope tracing experiments, namely the specific labeling of a single atom in a single feed compound.

1 Introduction

The insertion of stable isotopic tracers into living systems and subsequent tracking of isotopically labeled metabolites provides detailed insights into cellular and organismal metabolic activity [35]. The analysis of such data, however, is complicated by the complexity of metabolic networks, which usually harbour multiple alternative pathways generating compounds with varying efficiency [66]. Despite advances in the reconstruction of metabolic capabilities from genome annotations [49], models of metabolic networks are most likely incomplete. Moreover, not all metabolic reactions are necessarily well understood as far as their reaction mechanism is concerned. Isotope tracing experiments have the potential to correct, extend, and

improve models of metabolic networks and their constitutent chemical reactions. A first step towards this goal is the efficient and robust prediction of isotope labeling patterns throughout large reaction networks.

Rates of metabolic reactions usually vary over time. However, it is generally believed that in fixed environments, fluxes will converge towards a stable stationary state. In fact, steady-states fluxes lie at the heart of Metabolic Flux Analysis (MFA) [20] and Flux Balance Analysis (FBA) [46]. In this setting, isotope tracing experiments are completely described by temporal changes of the relative abundance of the possible labeling patterns of each metabolite.

A practical difficulty in the analysis of isotope labeling experiments is the exponentially large number of distict isotopomers of each molecular type, since, at least in principle, each individual atom may appear in both labeled and unlabeled form. The most commonly employed experimental methods, moreover, only distinguish the number but not the location of (usually heavy) isotopes in each compound. In order to model such mass isotopomer distributions (MID) [29], the notion of *elementary metabolite units* (EMU) was introduced in [6]. Despite resulting simplification of isotope-level reaction networks, however, computational efforts remain substantial. The literature, furthermore, appears to lack a concise formal treatment of the construction and structure of the resulting networks on EMU-level reactions, even though several software tools have been built upon this approach.

Position-specific isotope analysis (PSIA) has become feasible at high resolution and throughput with modern mass spectrometry methods such as Orbitrap that produce fragmentation spectra for each compound [22, 26]. NMR approaches to PSIA are available as well [31]. Such methods yield in particular position specific labeling fractions $\lambda(u)$ for each atom in chemical reaction networks (CRN). However, these data are far from routine in metabolomics applications. Hence, it is of practical interest to better understand the formal structure of atom transition networks (ATN) and their derivation from CRNs and fluxes on a CRN.

In this contribution we focus on the computation of transmission matrices that describe the symmetries and stoichiometries of each reaction in

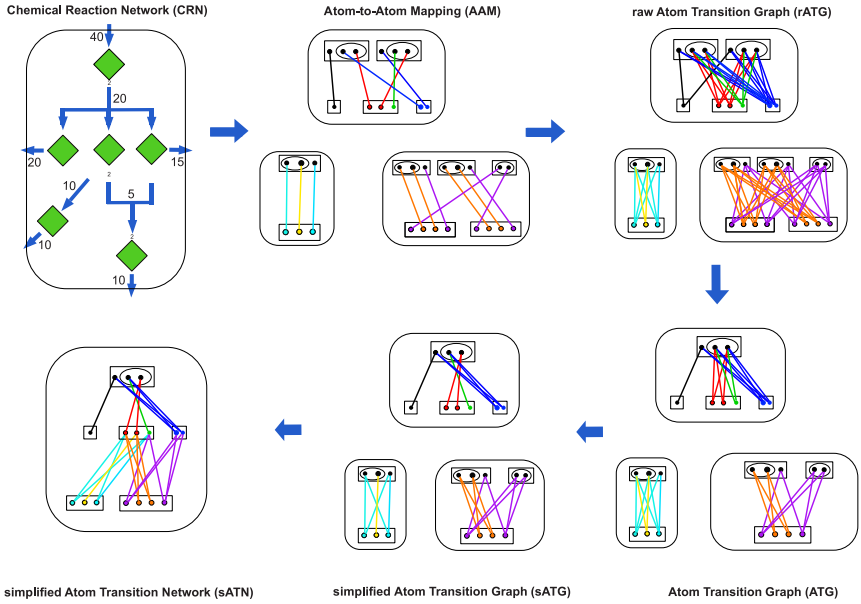

Figure 1. Overview of the formal constructions developed within this contribution. Atom-to-atom maps (AAM) derived from chemical reaction networks (CRN) with established steady-state fluxed are engaged for the construction of reaction-wise raw atom transition graphs (rATG). Transformation of rATG into atom transition graphs (ATG) account for the fact that atom transition networks (ATN) contain only one copy of each unique atom. Substantial reduction of complexity can be further obtained by the notion of simple atom transition graphs (sATG), which allow for the construction of a simplified ATN by the stepwise addition of reactions. Edge-weights are adjusted accordingly (see the respective subsections for details). For the CRN, compounds are depicted as diamonds with arrows indicating reactions connecting them. Fluxes are denoted as integer values beside reaction-edges, while stoichiometries other than 1 are added to the start of reaction edges. For the residual graphs, compounds are depicted as squares, atoms as circles, orbits of symmetric atoms are surrounded by ellipses in reactants, while marked in the same colour for products. Transition edges are naturally directed from reactants (top) to products (down) but arrow tips have been omitted for the sake of visibility. Feed and drain reservoirs are not shown explicitly.

full generality, and on their composition in a large CRN. We start from the atom-to-atom maps (AAM) of individual reactions in Sect. 3. The emphasis in this section is on the correct treatment of symmetries, including stoichiometric coefficients, in a single reaction. In addition to the AAM, only the orbits of the symmetry groups of each reactant and product molecules are needed. This results in weighted atom transition graphs (ATG). In addition we derive a dynamically equivalent sparser representation in Sect. 3.3 to further simplify the construction. We then investigate the concatenation of single reaction ATGs into networks, Sect. 4. A graphical overview of the entire construction is given in Fig. 1.

The complexity of metabolic networks entails numerous alternative pathways generating the same desired compound with varying efficiency [66]. The contributions of pathways is determined by fluxes, i.e., the rate of turnover of molecules through a pathway. Fluxes, however, are an emergent property of a metabolic system that cannot be measured directly because production and consumption of metabolites are usually superpositions determined by multiple reactions. In many situations, furthermore, metabolic systems can be assumed to be in or close to a steady-state. Flux estimates are nevertheless crucial for a great number of research goals relating to metabolic phenotypes and associated function and adaptations, including the modeling of metabolic diseases [11] and metabolic engineering [5]. Fluxes can be inferred in many cases from the structure of the CRN in conjunction with additional constraints, such as known feed sets and objective functions such as the biomass function. FBA [49] and MFA [20] are well-established methods for this task. It is very difficult, however, to verify the predicted fluxes experimentally.

Isotope labeling data can be used for this purpose. In the most general case, relative abundances of stable isotopes at specific positions of each metabolite are accessible at least in principle. The connection between fluxes and isotope labeling patterns, however, is based on a detailed mechanistic model that requires additional information. Three distinct tasks are of particular interest in this context:

(i) Given a flux, predict the expected pattern of isotope labels

(ii) Given experimental and predicted measurements, utilize their discrepancies to adjust estimated fluxes and possibly also the underlying network itself.

(iii) Find a good labeling strategy to distinguish with minimal experimental effort between competing models for a CRN and fluxes.

While it is possible to consider also temporal changes in metabolic fluxes [4], it is more common to consider isotope labeling data in a setting of constant fluxes. The majority of currently available methods nevertheless relies on solving large, complex systems of differential equations [67, 68, 69]. A major complication arises from correlations between isotope labels that are propagated together through a reaction. Predicting the MID, which is the quantity most easily accessible by mass spectrometry methods, therefore requires that one keeps track of individual isotopomers or at least all EMUs [6].

In contrast, position-specific label abundances are independent of correlations between atoms. The results of Sect. 3 and 4 in fact imply that they can be computed directly from fluxes and weights of the atom transition graphs. In Sect. 5 we use this fact to derive a simple dynamical system describing the propagation of atom-wise isotope labels through an entire CRN. Assuming that the underlying metabolic network already features steady fluxes, these can be traced using the temporal changes of relative, position-wise isotope abundances only. The dynamics of this process is described as a simple linear inhomogenous system of ordinary differential equations that can be solved explicitly. Moreover, the qualitative behaviour can be analyzed completely. We close with examples illustrating the impact of the formalism developed here.

2 Preliminaries

2.1 Notation

Graphs and Molecules A graph $G = (V, E)$ consists of a set of vertices V and a set of edges E such that every edge connects a pair of vertices,

i.e. $\{x, y\}$ and (x, y) for $x, y \in V$ in undirected and directed graphs, respectively. Throughout, we will consider vertex and edge labels defined as functions $\ell_V : V \to L_V$ and $\ell_E : E \to L_E$, where L_V and L_E are finite sets of labels. Molecules thus appear as connected undirected graphs with vertex labels representing atom types and edge labels representing bond types. We refer to [24] for a more extensive discussion of molecular graphs.

For directed graphs we distinguish the set of in-edges $E_{in}(v) := \{(u, v) \in E | u \in V\}$ and out-edges $E_{out}(v) := \{(v, u) \in E | u \in V\}$ at each vertex $v \in V$. Moreover, it will be useful to associate a weight: $w : E \to \mathbb{R}$ to each edge. The weighted out-degree is then $(D^{out})_{vv} := \sum_{e \in E_{out}(v)} w(e)$. Analogously, the weighted adjaceny matrix A^w has entries $A^w_{uv} := w((u, v))$ for $(u, v) \in E$ and $A^w_{uv} := 0$ otherwise. Denoting by D^{out} the diagonal matrix of weighted out-degrees, we define the Laplacian matrix of a weighted directed graph as

$$L = D^{out} - A^w \tag{1}$$

For each vertex v define the set of reachable vertices such that $z \in \mathfrak{R}_G(v)$ if there is a path from v to z. A vertex set W in an undirected graph G is connected if for all $x, y \in W : y \in \mathfrak{R}(x)$ and $x \in \mathfrak{R}(y)$. For directed graphs, W is called strongly connected if the same condition holds. W is weakly connected in a directed graph if it is connected in the undirected transformation of it ignoring edge direction. We call maximal sets W respectively the strongly connected, weakly connected, or connected components of G.

A bijective map $\mu : V(G) \to V(H)$ between two graphs G and H is an isomorphism if it preserves adjacency and labels. An automorphism of a graph G is a map $\varrho : V(G) \to V(G)$ that preserves adjacency and labels, i.e., an isomorphism of G to itself. The automorphisms of G form the group $\mathrm{Aut}(G)$ under composition. The set $\mathrm{orb}(x) := \{y \in V(G) | \exists \varrho \in \mathrm{Aut}(G) : \varrho(x) = y\}$ is called the orbit of x. Clearly, the orbits form a partition of $V(G)$.

Chemical reactions, directed hypergraphs and atom-to-atom maps In its most general form, a chemical reaction r can be understood as a transformation between multi-sets of molecules of the form

$$\sum_{c \in C} s_{cr}^{-} \cdot c \quad \longrightarrow \quad \sum_{c \in C} s_{cr}^{+} \cdot c \tag{2}$$

where the *stochiometric coefficients* $s_{cr}^{-} \geq 0$ and $s_{cr}^{+} \geq 0$ denote the multiplicity of molecule c among the reactants and products of reaction r, respectively. Two alternative representations of a reaction will be used here. First, we can interpret each reaction as a directed multi-hypergraph whose vertex set C consists of the molecules. Each edge corresponds to a reaction r and can be formally written as pair of multisets (r^{-}, r^{+}), where c is contained in r^{-} with multiplicities s_{cr}^{-} if $s_{cr}^{-} > 0$ and in r^{+} with multiplicities s_{cr}^{+} if $s_{cr}^{+} > 0$, respectively. Formally this can be written as $r^{\pm} = \sum_{c \in C} s_{cr}^{\pm} \cdot c$. These multisets are known as *complexes* in the literature on chemical reaction networks [23]. Each reaction r thus can be thought of as a transformation $r^{-} \to r^{+}$ of two complexes. Generalizing the notation for graphs we write $r \in E_{in}(c)$ if $c \in r^{+}$, i.e., if compound c is a product of reaction r and $r \in E_{out}(c)$ if $c \in r^{-}$, i.e., if compound c is a reactant in reaction r.

By definition, a complex $r^{\pm}$ is simply the disjoint union of the reactants or products, respectively. Each molecule, on the other hand, can itself be represented as a connected undirected graph. A complex $r^{\pm}$ thus has a representation as an undirected graph $Q_r^{\pm}$, with connected components that are the molecular graphs of the contributing chemical compounds. A key property of chemical reactions is the preservation of atoms.

Definition 1. An *tom-to-atom map (AAM)* for a reaction $r = (Q \to Q')$ is a bijection of the vertex sets of the complexes $\varphi : V(Q) \to V(Q')$ that preserves atom types and hence satisfies $\ell_{V(Q)}(x) = \ell_{V(Q')}(\varphi(x))$ for all $x \in V(Q)$.

The AAM of r provides an atom-level description of the chemical reaction, and thus of the bonds that change in the course of the reaction. For a detailed investigation of the structure of AAMs we refer to [27].

Reaction Networks and Fluxes In order to define flows on directed (multi)hypergraphs such as CRNs, first, we identify subsets $F, D \subseteq C$ of compounds that take on the role of food and drain. For these we add *transport reactions* of the form

$$\varnothing_c \longrightarrow c \qquad \text{and} \qquad c' \longrightarrow \varnothing_{c'} \tag{3}$$

and add the external feed and drain sets $C_F := \{\varnothing_c | c \in F\}$ and $C_D := \{\varnothing_{c'} | c' \in D\}$ to the network, setting $\tilde{C} := C \cup C_F \cup C_D$. These designate external reservoirs for the reaction network under consideration. The "pseudo-reactions" $(\varnothing_c \to c)$ and $(c' \to \varnothing_{c'})$ transport molecules of type $c \in F$ into the system and remove molecules of type $c' \in D$. We write R_F and R_D for these sets of *feeding* and *draining* reactions and set $\tilde{R} = R \cup R_F \cup R_D$. We note that in some cases it may be useful to model the concerted transport of multiple compounds. This can be handled, however, by inserting a formal reactant or product of such reactions together with import or export of the formal intermediate. We therefore do not sacrifice generality when considering import and export of individual compounds only. We will make explicit use of this simple form of the import and export later-on.

A (hyper)flow on a directed multi-hypergraph, possibly with both feeding and draining reactions, is a function $F : R \to \mathbb{R}_0^+$ such that for all $c \in C$ holds

$$\sum_{r \in \tilde{R}} s_{cr}^+ F(r) = \sum_{r \in \tilde{R}} s_{cr}^- F(r) \tag{4}$$

This condition is known as *flux balance*. In our discussion below it will be convenient to attach further formal feed and drain reactions that feed the set C_F and drain the set C_D. In this case, flux balance extends to $\tilde{C}$.

Every hyperflow describes a stationary state in the sense that for each compound c, the total production plus influx by feeding equals the total consumption plus outflux by draining. The notation simplifies considerably by introducing the stoichiometric matrix $\mathbf{S}$ with entries $\mathbf{S}_{cr} := s_{cr}^+ - s_{cr}^-$ for all $c \in C$ and all reactions $r \in R$, including the feeding and draining reactions. For simplicity, we do not include the extra vertices $\varnothing_s$ and $\varnothing_d$ in

the definition of $\mathbf{S}$. In vector form, the flow condition simplifies to $\mathbf{S}\mathit{F} = 0$.

A *flux* is a function $\mathit{F} : R \times [0, \infty]$ that measures how many instances of the complex r^- are transformed to r^+ per time unit at a given time t. In a deterministic setting, we have

$$\frac{dx}{dt} = \mathbf{S}\mathit{F}(t) \tag{5}$$

where x is the vector of concentrations x_c of compounds $c \in C$, see [39] for more details.

This expression reduces to $\mathbf{S}\mathit{F}(t) = 0$ in the steady-state, i.e., if the flux becomes independent of time. A steady-state flux, $\dot{\mathit{F}} = 0$, therefore satisfies $\mathbf{S}\mathit{F}(t) = 0$ and hence also flux balance. The terms (hyper)flow and steady-state flux, thus, are synonymous.

2.2 Flux decomposition theorem for chemical reaction networks

A well-known result for network flows is that every flow $f : R \to \mathbb{R}$ can be decomposed into constant flows along paths from source to drain and flows that are constant along directed cycles [25]. Moreover, the total out-flow of the source necessarily equals the total in-flow into the sink. Flows on hypergraphs [1, 13] in general do not satisfy analogous results. Similar statements, however, hold for *chemical* reaction networks. The properties discussed in this section seem to be known in the community, however, there does not appear to be a convenient reference. We therefore include proofs here for completeness.

We start by briefly considering properties of CRNs. A *cornucopia* in a reaction network (C, R) is a non-negative linear combination of reactions that yields a net production of some compound $c' \in C$ without compensating net consumption of building material. This situation is described by a function $g : R \to \mathbb{R}_0^+$ such that $\sum_{r \in R} \mathbf{S}_{cr} g(r) \geq 0$ for all $c \in C$, and there is at least one $c' \in C$ such that $\sum_{r \in R} \mathbf{S}_{c'r} g(r) > 0$. Clearly, cornucopias and *abysses*, their counterparts with $\sum_{r \in R} \mathbf{S}_{cr} g(r) \leq 0$ and $c' \in C$ such that $\sum_{r \in R} \mathbf{S}_{c'r} g(r) < 0$, cannot exist in CRNs since each reaction $r \in R$ conserves all participating atoms, and hence in particular molecular mass.

This property is captured abstractly by the existence of a strictly positive reaction invariant [32]:

Definition 2. Let $\mathbf{S}$ be the stoichiometric matrix of a reaction network (C, R). Then (C, R) is *conservative* if there is a function $\mu : C \to \mathbb{R}$ such that $\sum_{c \in C} \mu(c)\mathbf{S}_{cr} = 0$ and $\mu(c) > 0$ for all $c \in C$.

This condition is equivalent to ruling out the existence of abysses and cornucopias and ensures that the compounds $c \in C$ have a representation as sum formulas and structure formulas such that each reaction $r \in R$ preserves atoms [41].

Definition 3. A flow F of $(\tilde{C}, \tilde{R})$ is a *circulation* if $\mathit{F}(r) = 0$ for all $r \in R_F \cup R_D$.

Circulations are sometimes also called futile cycles [41]. The absence of abysses and cornucopias suggests that every flow is either a circulation or has positive values on both feed and drain reactions. This is indeed the case:

Lemma 1. *Let $(\tilde{C}, \tilde{R})$ be a conservative reaction network and let $\mathit{F} : \tilde{R} \to \mathbb{R}_0^+$ be a flow without drain, i.e., $\mathit{F}(r) = 0$ for all $r \in R_D$, or a flow without feed, i.e., $\mathit{F}(r) = 0$ for all $r \in R_F$. Then F is a circulation.*

Proof. Writing $g_\circ(r) \coloneqq \mathit{F}(r)$ for $r \in R$ and $g_\circ(r) \coloneqq 0$ for $r \in R_F$, as well as $g_F(r) \coloneqq \mathit{F}(r)$ for $r \in R_F$ and $g_F(r) \coloneqq 0$ for $r \in R$ we obtain $\mathit{F} = g_\circ + g_F$ since, by assumption, there is no draining reaction. Using that F is a flow, we have $\mathbf{S}\mathit{F} = \mathbf{S}g_\circ + \mathbf{S}g_F = 0$. Let C^+ be the set of species with an active feeding reaction, i.e., $[\mathbf{S}g_F]_c > 0$ for all $c \in C^+$ and $[\mathbf{S}g_F]_c = 0$ for all $c \in C \setminus C^+$. Therefore $[\mathbf{S}g_\circ]_c \leq 0$ for all $c \in C$ and $[\mathbf{S}g_\circ]_c < 0$ for $c \in C^+ \subseteq C$. If $C^+ \neq \emptyset$ then $g_\circ$ is an abyss in (C, R), which is not possible in conservative reaction networks. If $C^+ = \emptyset$ then there is no feeding reaction, i.e., $g_F = 0$, and thus $\mathit{F} = g_\circ$, which implies that F is a circulation. An analogous argument shows that in the absence of feeding reactions there also cannot be a draining reaction, and thus F is again a circulation. ∎

Moreover, if F is not a circulation, then we expect that F specifies a pathway, i.e., a connected sequence of reactions from feed to drain species. We next show that this intuition is also correct:

Definition 4. A flow F *connects from food to drain* if there is a sequence of reactions $r_0, r_1, \ldots, r_k, r_{k+1}$ with $F(r_i) > 0$, $r_0 \in R_F$, $r_{k+1} \in R_D$ such that a product of r_i is a reactant of r_{i+1}, for $0 \leq i \leq k$.

Lemma 2. *Let F be a flow on $(\tilde{C}, \tilde{R})$ then either F is a circulation or F connects from feed to drain.*

Proof. Denote by Γ_F the graph with vertex set $\tilde{C}$ and edges $(u, v) \in E(\Gamma_F)$ if there is a reaction r with reactant u and product v and $F(r) > 0$. Flow conservation immediately implies that only vertices in C_F and C_D can be source or sink vertices in Γ_F.

Let $\Gamma_F^{(i)}$ be a weakly connected component of Γ_F and set $F^{(i)}(r) = F(r)$ if all reactants and products of r are vertices of $\Gamma_F^{(i)}$. In particular, therefore, for each $c \in V(\Gamma_F^{(i)}) \cap C$ all in- and out-flows defined by F are contained in $F^{(i)}$ and thus $F^{(i)}$ is again a flow. We can therefore write $F = \sum_i F^{(i)}$ as sum of flows on the weakly connected components of Γ_F. In particular, therefore, each $F^{(i)}$ is either a circulation or, by Lemma 1, contains both a feed and a drain.

Let $F^{(i)}$ be a flow with source and drain vertices, and assume that $\Gamma_F^{(i)}$ does not contain a directed path between as source vertex and drain vertex. Then the vertex set $V(\Gamma_f^{(i)})$ contains disjoint subsets V^+ and V^-, such that V^+ contains all source vertices and V^- contains all sink vertices in $\Gamma_F^{(i)}$. In particular, therefore, there is no reaction that has a reactant in V^+ and a product in V^-, and *vice versa*, and thus there is no edge in $\Gamma_F^{(i)}$ between V^+ and V^-, contradicting the assumption that $\Gamma_F^{(i)}$ is a connected component. Thus $\Gamma_f^{(i)}$ must contain a directed path between a source and a sink vertex. ■

The proof of Lemma 2 immediately implies

Corollary. *Let F be a flow on $(\tilde{C}, \tilde{R})$. Then the restriction of F to every connected component of the graph Γ_F is either a circulation or a pathway connecting feed and drain.*

We say that a flow F *contains* the circulation $F_\circ$ if $F_\circ(r) \leq F(r)$ for all $r \in \tilde{R}$.

Lemma 3. *Let F be a flow on $(\tilde{C}, \tilde{R})$ and suppose there is a circulation $F_\circ$ on $(\tilde{C}, \tilde{R})$ such that $F_\circ(r) \leq F(r)$ for all $r \in \tilde{R}$. Then there is $\alpha > 0$ such that $F' = F - \alpha F_\circ$ is a flow on $(\tilde{C}, \tilde{R})$ and there is a reaction r^* such that $F(r^*) > 0$ and $F'(r^*) = 0$.*

Proof. If suffices to choose $\alpha = \min_{r:F_\circ(r)>0} F(r)/F_\circ(r)$. For the corresponding reaction $r^* = \arg\min_{r:F_\circ(r)>0} F(r)/F_\circ(r)$ we have $\alpha = F(r^*)/F_\circ(r^*)$ and thus $F'(r^*) = 0$. Moreover, one easily checks that we have $0 \leq \alpha F_\circ(r) \leq F(r)$ for all $r \in \tilde{R}$: and thus $F'(r) = F(r) - \alpha F_\circ(r) \geq 0$. Since both F and $F_\circ$ satisfy flow conservation for all $c \in C$, this is also true for $F' = F - \alpha F_\circ$. Thus F' is again a flow on $(\tilde{C}, \tilde{R})$. ■

Since F' has at least one active reaction less than F, we conclude that we can remove no more than $|R|$ circulations from F until we obtain a flow F^* on $(\tilde{C}, \tilde{R})$ that does not contain any further circulations. Note that the sum of circulations is again a circulation. Therefore we have:

Proposition 1. *Every flow F on $(\tilde{C}, \tilde{R})$ can be decomposed into a circulation $F_\circ$ and a circulation-free-flow $F^\perp$ that contains both feed and drain reactions.*

It is worth noting, however, that this decomposition is not unique. In general: the same flow can be written in different ways as a non-negative linear combination of elementary flux modes [53, 70].

For flows on graphs, the total feed equals the total drain. In chemical networks, however, particle numbers change and hence stoichiometric coefficients different from unity appear in Equ. (4). As a consequence, the balance of feed and drain also takes a somewhat more complicated form:

Lemma 4. *Let F be a flow on a reaction network $(\tilde{C}, \tilde{R})$. Then every reaction invariant $\mu : \tilde{C} \to \mathbb{R}$ satisfies*

$$\sum_{c \in C_F} \mu(c) F(r_c) = \sum_{c' \in C_D} \mu(c') F(r_{c'}) \tag{6}$$

where $r_c = (\varnothing_c \to c) \in R_F$ and $r_{c'} = (c' \to \varnothing_{c'}) \in R_D$.

Proof. Let μ be a reaction invariant, i.e., $\sum_{c\in\tilde{C}} \mu(c)\mathbf{S}_{cr} = 0$. Using that the stoichiometric coefficients for the feed and drain reactions are 1 for the external version of the feed and drain molecules, we obtain:

$$\sum_{r\in\tilde{R}} \underbrace{\sum_{c\in\tilde{C}} \mu(c)\mathbf{S}_{cr}}_{=0} f(r) = \sum_{c\in C_F} \mu(c) \underbrace{(-s^-_{cr_c})}_{-1} F(r_c) + \sum_{c'\in C_D} \mu(c') \underbrace{s^+_{c'r_{c'}}}_{+1} F(r_{c'})$$
$$+ \sum_{c\in C} \mu(c) \underbrace{\sum_{r\in\tilde{R}} \mathbf{S}_{cr} F(r)}_{=0}$$

Thus both the l.h.s. and the sum over the inner species vanishes and we are left with the statement of the lemma. ■

For flows on graphs there is no distinction between different compounds, hence the identity $\mu(c) = 1$ is always a reaction invariant. In chemical network, $\mu(c)$ is e.g. the number of any type of atom in compound c or the molecular mass.

3 The atom transition graph of a reaction

3.1 Atom-to-atom maps

Consider a chemical reaction network $(\tilde{C}, \tilde{R})$ with feed and drain reactions. Each reaction $r = (Q_r \to Q'_r)$ is described by an AAM $\varphi_r : V(Q_r) \to V(Q'_r)$ that determines the fate of each atom during the reaction. We note in passing that the assumption of an AAM immediately implies that $(\tilde{C}, \tilde{R})$ is conservative since $|V(c)|$, the number of atoms in each compound, is a positive reaction invariant. In principle, the fate of an atom $u \in V(c)$ in some compound $c \in \tilde{C}$ through a sequence of reactions specified as a flux F can be traced by "concatenating" the AAMs φ_r of the reactions r that are active in F, i.e., for which we have $F(r) > 0$. In practice, however, this "concatenation" is complicated by the fact that (i) given F, multiple reactions may use and/or produce the same compound, and, (ii) symmetries in the reactions need to be taken into account.

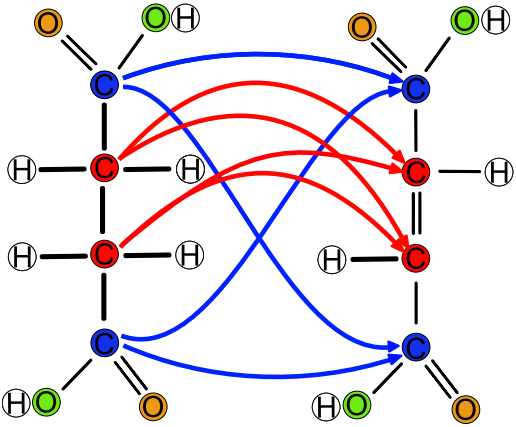

Figure 2. Illustration of the oxidation of succinate to fumarate in terms of carbon atom transition. Carbon and oxygen atoms of the same equivalent class, for each molecule seperately, are depicted in the same colour. Atom transitions edges for carbons are depicted in the colour of their appurtenant equivalence class.

We start with the latter issue, which can be understood by considering a single reaction $r = (Q \to Q')$ with AAM φ in isolation. Note that we drop the explict reference to r throughout this section. Let $\varrho : V(Q) \to V(Q)$ and $\varrho' : V(Q') \to V(Q')$ be automorphisms of the reactant and product graphs Q and Q', respectively. Atoms in the same orbit $\mathrm{orb}_Q(\,.\,)$ or $\mathrm{orb}_{Q'}(\,.\,)$ of $\mathrm{Aut}(Q)$ and $\mathrm{Aut}(Q')$, respectively, are chemically indistinguishable. This gives rise to a notion of equivalence of AAMs [33].

Definition 5. Let $\varphi : V(Q) \to V(Q')$ and $\psi : V(Q) \to V(Q')$ be two vertex label preserving bijections. Then φ and ψ are equivalent if there are automorphisms $\varrho \in \mathrm{Aut}(Q)$ and $\varrho' \in \mathrm{Aut}(Q')$ such that $\psi = \varrho' \circ \varphi \circ \varrho^{-1}$.

Equivalent AAMs describe *the same* chemical reaction. Fix an isotope labeled atom $x \in V(Q)$. Then the assignments $x \mapsto \varphi(x)$ and $x \mapsto (\varrho' \circ \varphi \circ \varrho^{-1})(x) = \varrho'(\varphi(\varrho^{-1}(x)))$ are chemically indistinguishable. Tracing of atoms across multiple reactions therefore must keep track of this symmetry-based ambiguity, see Fig. 2.

Definition 6. The *raw (reaction-wise) atom transition graph* (rATG) $\tilde{T}_{QQ'}$ of a single reaction $Q \longrightarrow Q'$ with AAM φ is the bipartite graph

with vertex set $V(T_{QQ'}) = V(Q) \uplus V(Q')$ and a set of directed edges $E(\tilde{T}_{QQ'}) = \bigcup_{x \in V(Q)} E_{out}(x)$ where $E_{out}(x) \coloneqq \{(x, \varrho'(\varphi(\varrho^{-1}(x))))|\varrho \in \mathrm{Aut}(Q), \varrho' \in \mathrm{Aut}(Q')\}$.

The edge sets $E_{out}(x)$ are disjoint by construction and correspond to chemically indistinguishable atom transitions. Fig. 2 shows the equivalence classes of AAM edges for the carbon atoms in the oxidation of succinate to fumarate. Fig. 3 illustrates the general relationship between a reactant and a product molecule implied by an AAM.

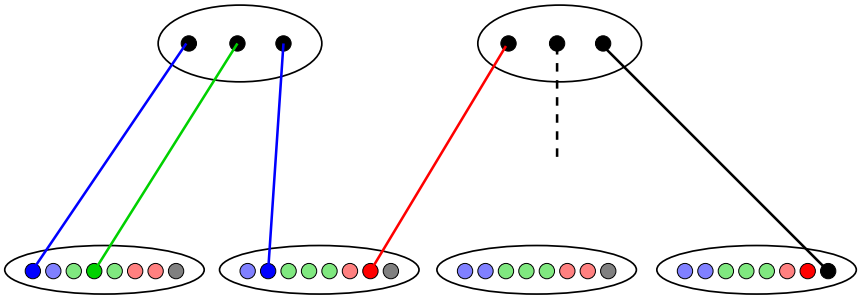

Figure 3. Construction of the raw Atom Transition Graph from an atom-to-atom map. Consider a reaction $r = (Q \to Q)'$ with 2 copies of $c \subseteq Q$ (top) as reactants and 4 copies of $c' \in Q'$ as product (below). We consider only the orbit $\mathrm{orb}_Q(x)$ of a fixed vertex x in Q and focus on a single product molecule c'. In this example, $|\mathrm{orb}_c(x)| = 3$ and thus $|\mathrm{orb}_Q(x)| = s_{cr}^{-} \cdot |\mathrm{orb}_c(x)| = 2 \times 3 = 6$. Of these 6 atoms, 5 are mapped to copies of c' by the AAM φ, and one is mapped to a different product. In c', four orbits $\mathrm{Aut}(c')$ comprising a total of 8 distinct atoms are reached from $\mathrm{orb}_c(x)$. Counting the number of atoms from $\mathrm{orb}_Q(x)$ transmitted to the orbits of $\mathrm{Aut}(Q')$ we obtain $\eta(\mathrm{black}, x) = 1$, $\eta(\mathrm{red}, x) = 1$, $\eta(\mathrm{green}, x) = 1$, $\eta(\mathrm{blue}, x) = 2$. The atom mapped to a different product molecule is not represented here.

3.2 Atom transition graph

It is not possible to compose the ATG of different reactions in the usual sense, however, since in general both reactant and product complex will be different for different reactions. The composition of reactions thus requires a decomposition of the complexes into their constituent chemical

compounds, i.e., the non-isomorphic connected components of Q and Q', respectively. For a reaction $r = (Q \to Q')$ we therefore introduce the $Q_\circ$ and $Q'_\circ$ as the disjoint union of the pairwisely non-isomorphic connected components Q and Q', respectively. We may think of $Q_\circ$ and $Q'_\circ$ as the sets of reactant and product molecules. By a slight abuse of notation we write $c \subseteq Q_\circ$ and $c' \subseteq Q'_\circ$ for a reactant and product graph, respectively. Note that we treat connected components of Q and Q' as different even if they are isomorphic, i.e., if the reaction r contains explicit catalysts. In the following we will write x and y for vertices (atoms) in the complexes Q and Q', respectively, while at the level of molecules we will use u and v.

Definition 7. $\zeta : Q \dot\cup Q' \to Q_\circ \dot\cup Q'_\circ$ is a map such that (i) $\zeta(Q) = Q_\circ$ and $\zeta(Q') = Q'_\circ$ and (ii) $\zeta(c)$ is a connected component of $Q_\circ$ or $Q'_\circ$ if and only if c is a connected component of Q or Q', and (iii) ζ is an isomorphism between connected components.

In particular, $\zeta^{-1}(u)$ for some vertex $u \in V(c)$ with $c \subseteq Q_\circ$ is an equivalent class of the vertices in Q. Analogously, $\zeta^{-1}(v)$ for some vertex $v \in V(c')$ with $c' \subseteq Q'_\circ$ is an equivalent class of the vertices in Q'. The equivalence classes $\zeta^{-1}(\,.\,)$ are sets of vertices in pairwise distinct isomorphic connected components of Q or Q'. Their size is therefore determined by the stochiometric coefficients: we have $|\zeta^{-1}(u)| = s^-_{cr}$ for all $u \in V(c)$ with $c \subseteq Q$ and $|\zeta^{-1}(v)| = s^+_{c'r}$ for $v \in V(c')$ with $c' \subseteq Q'$.

The map ζ also relates symmetries of Q or Q' with the symmetries of the constituent connected components: For $x \in V(Q)$ and $y \in V(Q')$ we have $|\operatorname{orb}_Q(x)| = s^-_{cr} \cdot |\operatorname{orb}_c(\zeta(x))|$ and $|\operatorname{orb}_{Q'}(y)| = s^+_{c'r} \cdot |\operatorname{orb}_{c'}(\zeta(y))|$, where $c \subseteq Q_\circ$ and $c' \subseteq Q'_\circ$ such that $\zeta(x) \in V(c)$ and $\zeta(y) \in V(c')$.

Definition 8. The atom transition graph (ATG) $T_r = T_{QQ'}$ of a reaction $r = (Q \to Q')$ is obtained as the quotient of the raw atom transition graph $\tilde T_{QQ'}$ w.r.t. the equivalence classes $\zeta^{-1}(\,.\,)$ defined by the isomorphic connected components of Q and Q', respectively.

Atoms are chemically indistinguishable w.r.t. to a given reaction r as they belong to the same orbit of automorphism groups of the reactant complex Q or the product complex Q'. Therefore we have to consider the

transmission of atoms not in terms of a fixed AAM $\varphi : V(Q) \to V(Q')$ but in terms of the orbits of $\mathrm{Aut}(Q)$ and $\mathrm{Aut}(Q')$, respectively.

Definition 9. Let $r = (Q \to Q')$ be a reaction, $x \in V(Q)$ and $y \in V(Q')$. Then $\eta(y, x)$ is the number of pairs (x', y') such that $x' \in \mathrm{orb}_Q(x)$, $y' \in \mathrm{orb}_{Q'}(y)$ and $y' = \varphi(x')$, i.e., (x', y') is a directed edge in the ATG of $Q \to Q'$.

In other words, $\eta(y, x)$ is the number of atoms in $\mathrm{orb}_Q(x)$ that are mapped by φ to an atom in $\mathrm{orb}_{Q'}(y)$. The order of the arguments of η will be convenient later-on for use in matrix notation. In particular, $\eta(y, x) = 0$ if the AAM does not transmit an atom from $\mathrm{orb}_c(x)$ to $\mathrm{orb}_{c'}(y)$. Moreover, since the AAM φ is a bijection, we have $\eta(y, x) \leq \min\{|\,\mathrm{orb}_Q(x)|, |\,\mathrm{orb}_{Q'}(y)|\}$. By construction $\eta(y, x) = \eta(y', x')$ for $x' \in \mathrm{orb}_Q(x)$ and $y' \in \mathrm{orb}_Q(y)$, i.e., $\eta(\,.\,,\,.\,)$ is really just a function of the orbits. Writing $\Theta(Q)$ and $\Theta(Q')$ for the set of orbits of $\mathrm{Aut}(Q)$ and $\mathrm{Aut}(Q')$, respectively, and $\mathbb{I}[\,.\,]$ for the indicator function with values $\mathbb{I}[\mathrm{true}] = 1$ and $\mathbb{I}[\mathrm{false}] = 0$, we therefore have

$$\eta(y, x) = \eta(Y, A) = \sum_{\bar{x} \in A, \bar{y} \in Y} \mathbb{I}[\bar{y} = \varphi(\bar{x})] \tag{7}$$

for $A \in \Theta(Q)$ and $Y \in \Theta(Q')$ such that $x \in A$ and $y \in Y$. Moreover, since $\zeta^{-1}(u) \subseteq \mathrm{orb}_Q(x)$ for all $x \in \zeta^{-1}(u)$, we may define $\eta(v, u) \coloneqq \eta(y, x)$ for any $x \in \zeta^{-1}(u)$ and $y \in \zeta^{-1}(v)$ as well as $\eta(Y_\circ, A_\circ) = \eta(v, u)$ for any $u \in A_\circ$ and $v \in Y_\circ$ with $A_\circ \in \Theta(Q_\circ)$ and $Y_\circ \in \Theta(Q'_\circ)$.

Lemma 5. *Let $r = (Q \to Q')$ be a reaction. Then there is a directed edge (x, y) in the raw ATG $\tilde{T}_{QQ'}$ if and only if $\eta(y, x) > 0$.*

Proof. By construction of $\tilde{T}_{QQ'}$, there is an edge (x, y) if and only if there is an edge (x', y') for all $x' \in \mathrm{orb}_Q(x)$ and $y' \in \mathrm{orb}_{Q'}(y)$. An edge is present if there is $x' \in \mathrm{orb}_Q(x)$ and $y' \in \mathrm{orb}_{Q'}(y)$ such that $y' = \varphi(x')$. The same condition defines $\eta(y, x) > 0$ in Equ. (7). ∎

Since φ is a bijection, there are exactly $|\,\mathrm{orb}_Q(x)|$ vertices that are the image of $\mathrm{orb}_Q(x)$ under φ, and each of them is contained in exactly one

orbit $Y \in \Theta(Q')$. Analogously, every $y' \in \mathrm{orb}_{Q'}(y)$ is the image of exactly one $x \in V(Q)$. Therefore we have

$$\sum_{Y \in \Theta(Q')} \eta(Y, A) = |A| \qquad \text{and} \qquad \sum_{A \in \Theta(Q)} \eta(Y, A) = |Y| \tag{8}$$

A *valuation* $\lambda : V(Q) \to \mathbb{R}$ is an arbitrary function that is preserved throughout a reaction in the sense that the value of $\lambda(x)$ is distributed uniformly over all $\mathrm{orb}_{Q'}(\varphi(x))$ and each $y \in V(Q')$ receives a uniform share of all $x \in \mathrm{orb}_Q(\varphi^{-1}(y))$. The valuation $\lambda' : V(Q') \to \mathbb{R}$ therefore can be written in the from

$$\begin{aligned} \lambda'(y) &\coloneqq \sum_{A \in \Theta(Q)} \frac{1}{|A|} \sum_{x \in A} \frac{1}{|Y|} \eta(y, x) \cdot \lambda(x) \\ &= \sum_{x \in V(Q)} \frac{\eta(y, x)}{|\,\mathrm{orb}_Q(x)| \cdot |\,\mathrm{orb}_{Q'}(y)|} \lambda(x) \end{aligned} \tag{9}$$

Since $\eta(y, x)$ is constant on $A \times Y$, we infer that $\lambda'(y)$ is constant on Y.

It remains to reduce these expressions to a molecule-wise representation by reducing the expressions to orbits on the individual molecules. The key observation is that $A \in \Theta(Q)$ and $Y \in \Theta(Q')$ are partitioned further by $\zeta(\,.\,)$ in the form

$$A = \bigcup\!\!\!\cdot_{u \in A_\circ} \zeta^{-1}(u), \quad \text{and} \quad Y = \bigcup\!\!\!\cdot_{v \in Y_\circ} \zeta^{-1}(v), \tag{10}$$

where $A_\circ = \zeta(A)$ and $Y_\circ = \zeta(Y)$. Recall that $|\zeta^{-1}(u)| = s^-_{c_u r}$ and $|\zeta^{-1}(v)| = s^+_{c'_v r}$, where c_u and c'_v are the molecules in which the atoms u and v are located.

First note for $v = \zeta(y)$ we have $\lambda'(v) = \lambda'(Y_\circ) = \lambda'(Y) = \lambda'(y)$, where by a slight abuse of notation we re-use λ and λ' for the valuations of the disjoint unions of non-isomorphic connected components. A short

computation now yields

$$\begin{aligned}\lambda'(v) = \lambda'(y) &= \sum_{A_\circ \in \Theta(Q_\circ)} \frac{1}{|A_\circ| \cdot s^-_{cr}} \sum_{u \in A_\circ} \sum_{x \in \zeta^{-1}(u)} \frac{1}{|Y_\circ| \cdot s^+_{c'_v r}} \cdot \eta(y,x) \cdot \lambda(x) \\ &= \sum_{A_\circ \in \Theta(Q_\circ)} \frac{1}{|A_\circ|} \sum_{u \in A_\circ} \frac{1}{s^-_{c_u r}} \cdot s^-_{c_u r} \cdot \frac{1}{|Y_\circ| \cdot s^+_{c'_v r}} \cdot \eta(v,u) \cdot \lambda(u) \\ &= \sum_{u \in V(Q_\circ)} \frac{1}{|\operatorname{orb}_{c'_v}(v)| \cdot s^+_{c'_v r}} \cdot \frac{1}{|\operatorname{orb}_{c_u}(u)|} \cdot \eta(v,u) \cdot \lambda(u)\end{aligned} \tag{11}$$

For each $u' \in \operatorname{orb}_c(u)$ and $v' \in \operatorname{orb}_{c'}(v)$ we therefore have a weight

$$h(v,u) \coloneqq \frac{\eta(v,u)}{s^+_{c'_v r} \cdot |\operatorname{orb}_{c'_v}(v)| \cdot |\operatorname{orb}_{c_u}(u)|} \tag{12}$$

As an immediate consequence we can rewrite Equ.(9) in the following simple form

$$\lambda'(v) = \sum_{u \in V(Q_\circ)} h(v,u)\, \lambda(u)\,. \tag{13}$$

Not surprisingly, the weights $h(\,.\,,\,.\,)$ are closely related to the structure of the ATG:

Lemma 6. *The ATG $T_{QQ'}$ contains the edge (u,v) if and only if $h(v,u) > 0$.*

Proof. We have $h(v,u) > 0$ on $V(Q_\circ) \uplus V(Q'_\circ)$ whenever there are $u' \in \operatorname{orb}_{c_u}(u)$ and $x \in \zeta^{-1}(u')$ such that $\zeta(\varphi(x)) \in orb_{c'_v}(v)$. In other words, there is a directed edge from u to v if and only if there is a directed edge between $\operatorname{orb}_Q(x)$ and $\operatorname{orb}_{Q'}(y)$ in $\tilde{T}_{QQ'}$ and thus, if and only if there is a directed edge in the ATG $T_{QQ'}$. ∎

Lemma 7. *For every reaction $r = (Q \to Q')$ we have $\sum_{u \in V(Q_\circ)} h(v,u) = 1$ and for all $v \in Q'_\circ$, and for all $u \in Q_\circ$ it holds*

$$\sum_{c' \subseteq Q'_\circ} s^+_{c'r} \sum_{v \in V(c')} h(v,u) = s^-_{c_u r}\,. \tag{14}$$

Proof. Recalling that $\eta(\,.\,,\,.\,)$ is constant on orbits we can write $\eta(v, A_\circ) \coloneqq \eta(Y_\circ, A_\circ)$ with $Y_\circ = \mathrm{orb}_{Q_\circ}(v)$ and $v \in Y_\circ$ and $\eta(v, A) \coloneqq \eta(Y, A)$ with $Y = \mathrm{orb}_Q(y)$ for some $y \in \zeta^{-1}(v)$. Using Equ.(12) we then obtain

$$\begin{aligned}
\sum_{u\in V(Q_\circ)} h(v,u) &= \sum_{c\subseteq Q_\circ} \sum_{A_\circ\in\Theta(c)} \sum_{u\in A_\circ} \frac{\eta(v,u)}{s^+_{c'_v r}\cdot|\,\mathrm{orb}_{c'_v}(v)|\cdot|\,\mathrm{orb}_{c_u}(u)|} \\
&= \frac{1}{s^+_{c'_v r}\cdot|\,\mathrm{orb}_{c'_v}(v)|} \sum_{c\subseteq Q_\circ} \sum_{A_\circ\in\Theta(c)} \eta(v,A_\circ) \sum_{u\in A_\circ} \frac{1}{|A_\circ|} \\
&= \frac{1}{s^+_{c'_v r}} \cdot \frac{1}{|\,\mathrm{orb}_{c'_v}(v)|} \sum_{A_\circ\in\Theta(Q_\circ)} \eta(v,A_\circ) \\
&= \frac{1}{s^+_{c'_v r}} \cdot \frac{1}{|\,\mathrm{orb}_{c'_v}(v)|} \sum_{A\in\Theta(Q)} \eta(v,A) \\
&\overset{Equ.\ (8)}{=} \quad \frac{1}{s^+_{c'_v r}} \cdot s^+_{c'_v r} = 1
\end{aligned}$$

Substituting Equ. (12) into the l.h.s. of Equ.(14) and using an analogous short hand notation yields

$$\begin{aligned}
&\sum_{c'\subseteq Q'_\circ} s^+_{c' r} \sum_{Y_\circ\in\Theta(c')} \sum_{v\in Y_\circ} \frac{\eta(v,u)}{s^+_{c' r}\cdot|\,\mathrm{orb}_{c'_v}(v)|\cdot|\,\mathrm{orb}_{c_u}(u)|} \\
&\quad= \frac{1}{|\,\mathrm{orb}_{c_u}(u)|} \sum_{c'\subseteq Q'_\circ} s^+_{c' r} \frac{1}{s^+_{c' r}} \sum_{Y_\circ\in\Theta(c')} \eta(Y_\circ,u) \underbrace{\sum_{v\in Y_\circ} \frac{1}{|Y_\circ|}}_{=1} \\
&\quad= \frac{1}{|\,\mathrm{orb}_{c_u}(u)|} \sum_{c'\subseteq Q'_\circ} \sum_{Y_\circ\in\Theta(c')} \eta(Y_\circ,u) \\
&\quad= \frac{1}{|\,\mathrm{orb}_{c_u}(u)|} \sum_{Y_\circ\in\Theta(Q'_\circ)} \eta(Y_\circ,u) \\
&\quad= \frac{1}{|\,\mathrm{orb}_{c_u}(u)|} \sum_{Y\in\Theta(Q)} \eta(Y,u) \quad \overset{Equ.\ (8)}{=} \quad s^-_{c_u r}
\end{aligned}$$

■

Valuations and their propagation across a reaction give rise to conserved quantities, i.e., reaction invariants of the form

Lemma 8. *Let $r = (Q \to Q')$ be a reaction, $\Lambda(c) \coloneqq \sum_{u \in V(c)} \lambda(u)$ for all $c \subseteq Q_\circ$, and $\Lambda'(c') \coloneqq \sum_{v \in V(c')} \lambda'(v)$ for all $c' \subseteq Q'_\circ$. Then*

$$\sum_{c \subseteq Q_\circ} \Lambda(c) \cdot s^-_{cr} = \sum_{c' \subseteq Q'_\circ} \Lambda'(c') \cdot s^+_{c'r} \tag{15}$$

Proof. Equ. (13) yields $\sum_{v \in V(c')} \lambda'(v) = \sum_{u \in V(Q_\circ)} \sum_{v \in V(c')} h(v,u) \cdot \lambda(u)$. This implies:

$$\begin{aligned}
\sum_{c' \subseteq Q'_\circ} \Lambda(c') \cdot s^+_{c'r} &= \sum_{c' \subseteq Q'_\circ} \sum_{u \in V(Q_\circ)} \sum_{v \in V(c')} h(v,u) \cdot \lambda(u) \cdot s^+_{c'r} \\
&= \sum_{u \in V(Q_\circ)} \underbrace{\left(\sum_{c' \subseteq Q'_\circ} s^+_{c'r} \sum_{v \in V(c')} h(v,u) \right)}_{s^-_{c_u r} \quad \text{by Equ. (14)}} \lambda(u) \\
&= \sum_{c \subseteq Q_\circ} s^-_{cr} \sum_{u \in V(c)} \lambda(u) = \sum_{c \subseteq Q_\circ} s^-_{cr} \cdot \Lambda(c)
\end{aligned}$$

■

We can therefore view the valuations $\lambda(\,.\,)$ as "building blocks" of chemically relevant invariants describing the transmission of the atom-level properties. For example, if we set $\lambda(u) = 1$ for all atoms of given type, then $\Lambda(c)$ measures the number of atoms of this type in compounds c and Equ.(15) ensures that this is indeed a reaction invariant.

3.3 Simplified atom transition graphs

In the following, we will be interested in particular in valuations that are constant on $\mathrm{orb}_{c_u}(u)$ for all $c \subseteq Q$ and all $u \in V(c)$. We shall see that in this case we can simplify the construction of the ATN. In particular, the averaging over the $x' \in \mathrm{orb}_Q(x)$, respectively can be omitted. In this case, we can rewrite Equ.(9) in a different form that makes it possible to omit some of the edges in the ATG.

Lemma 9. *Let $r = (Q \to Q')$ be a reaction and $\lambda : V(Q) \to \mathbb{R}$ a valuation on Q that is constant on the orbits of* $\mathrm{Aut}(Q)$*, then*

$$\lambda'(Y) = \sum_{A \in \Theta(Q)} \frac{1}{|Y|} \sum_{y' \in Y} \sum_{x \in A} \lambda(x) \mathbb{I}[y' = \varphi(x)] \tag{16}$$

Proof. Since λ and λ' are constant on the orbits on Q and Q', we can rewrite Equ. (9) in the form $\lambda'(y) = \lambda'(Y) = \sum_{A \in \Theta(Q)} \lambda(A) \frac{1}{|Y|} \eta(Y, A)$. Substituting Equ.(7) for $\eta(Y, A)$, observing that $\lambda(A) = \frac{1}{|A|} \sum_{x' \in A} \lambda(x') = \lambda(x)$ for all $x \in A$, and exchanging the order of the two sums yields Equ.(16). ∎

Since the assumption that $\lambda(x)$ is contant on $\mathrm{orb}_Q(x)$ also implies that $\lambda(u)$ is constant on $\mathrm{orb}_{c_u}(u)$, we can use Equ.(16) to compute $\lambda'(v)$ as follows:

$$\begin{aligned}
\lambda'(v) &= \sum_{A \in \Theta(Q)} \frac{1}{|Y|} \sum_{x \in A} \sum_{\bar{y} \in Y} \mathbb{I}[\bar{y} = \varphi(x)] \cdot \lambda(x) \\
&= \sum_{A_\circ \in \Theta(Q_\circ)} \sum_{u \in A_\circ} \sum_{x \in \zeta^{-1}(u)} \frac{1}{|Y_\circ|} \cdot \frac{1}{s^+_{c'_v r}} \sum_{\bar{v} \in Y_\circ} \sum_{\bar{y} \in \zeta^{-1}(\bar{v})} \mathbb{I}[\bar{y} = \varphi(x)] \cdot \lambda(u) \\
&= \sum_{u \in V(Q_\circ)} \underbrace{\frac{1}{s^+_{c'_v r} |\,\mathrm{orb}_{c'_v}(v)|} \sum_{\bar{v} \in \mathrm{orb}_{c'_v}(v)} \sum_{x \in \zeta^{-1}(u)} \sum_{\bar{y} \in \zeta^{-1}(\bar{v})} \mathbb{I}[\bar{y} = \varphi(x)]}_{=: h^*(v,u)} \cdot \lambda(u)
\end{aligned} \tag{17}$$

We next give a direct combinatorial interpretation of $h^*(v, u)$. To this end we consider the subset of atoms in the equivalence class $\zeta^{-1}(u)$ of $V(Q)$ with an AAM image in $\mathrm{orb}_{Q'}(y)$:

$$\zeta_v^{-1}(u) \coloneqq \left\{ x' \in \zeta^{-1}(u) \middle| \exists y' \in \zeta^{-1}(v) : \varphi(x') \in \mathrm{orb}_{Q'}(y') \right\}, \tag{18}$$

Using $\mathrm{orb}_{Q'}(y') = \bigcup_{\bar{v} \in \mathrm{orb}_{c'_v}(v)} \zeta^{-1}(\bar{v})$ we observe that $\varphi(x')$ may be located in any of the equivalence classes. Thus we can rewrite this definition as

$$\zeta_v^{-1}(u) = \left\{ x' \in \zeta^{-1}(u) \middle| \exists \bar{v} \in \mathrm{orb}_{c'_v}(v) \colon \varphi(x') \in \zeta^{-1}(\bar{v}) \right\} \tag{19}$$

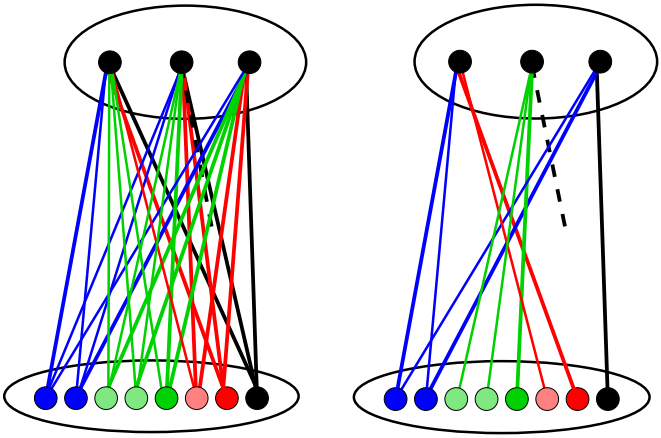

Figure 4. Comparison of the atom transition graph (ATG, left) and the simplified atom transition graph (sATG, right) of the example given in figure 3. Note that, by construction, the subgraph of $T_{QQ'}$ induced by the vertex set $\mathrm{orb}_c(u) \cup \mathrm{orb}_{c'}(v)$ is a complete bipartite graph, if and only if for any $x \in \zeta^{-1}(u), y \in \zeta^{-1}(v)$ there is an edge $(x, y) \in E(\tilde{T}_{QQ'})$.

Since for each x' there is at most one y' with $y' = \varphi(x')$ we can express the cardinality of $\zeta_v^{-1}(u)$ as follows:

$$|\zeta_v^{-1}(u)| = \sum_{x' \in \zeta^{-1}(u)} \sum_{\bar{v} \in \mathrm{orb}_{c'_v}(v)} \sum_{y' \in \zeta^{-1}(\bar{v})} \mathbb{I}[y' = \varphi(x')] \tag{20}$$

Comparing this expression with Equ. (17), we obtain

$$h^*(v, u) = \frac{1}{s^+_{c'_v r} \cdot |\,\mathrm{orb}_{c'_v}(v)|} |\zeta_v^{-1}(u)| \tag{21}$$

The propagation of valuations thus can also be expressed in the following form:

Corollary. *Let $r = (Q \to Q')$ be a reaction and assume that the valuation λ on $V(Q_\circ)$ is constant on the orbits of c for all $c \subseteq Q_\circ$. Then*

$$\lambda'(v) = \sum_{u \in V(Q_\circ)} h^*(v, u) \lambda(u)$$

This representation suggests to consider a simplified version of the ATG.

Definition 10. Let $r = (Q \to Q')$ be a reaction. Then the *simplified atom transition graph* $S_{QQ'}$ has vertex set $V(Q_\circ) \dot\cup V(Q'_\circ)$ and edges (u, v) if and only if $\zeta_v^{-1}(u) \neq \emptyset$.

Each edge (u, v) of $S_{QQ'}$ is associated with the weights $h^*(v, u)$. An illustration of the example from figure 3 is shown in figure 4. An alternative derivation of $S_{QQ'}$ and the edge weights $h^*(\,.\,,\,.\,)$ is given in the appendix.

In contrast to $h(v, u)$ defined in Equ.(12), the weights $h^*(v, u)$ are not necessarily constant on the $\mathrm{orb}_{c_u}(u)$ and $\mathrm{orb}_{c'_v}(v)$ due to the factor $|\zeta_v^{-1}(u)|$. However, the sum of the weights must be the same for all orbits $\mathrm{orb}_{c_u}(u)$ on the reactants, i.e.,

$$\sum_{u' \in \mathrm{orb}_{c_u}(u)} h^*(v, u') = \sum_{u' \in \mathrm{orb}_{c_u}(u)} h(v, u') = h(v, u) \cdot |\,\mathrm{orb}_{c_u}(u)|\,, \qquad (22)$$

where the last equality follows from the fact that $h(v, u)$ is constant on $\mathrm{orb}_{c_u}(u)$. A short computation then yields

$$\eta(v, u) = \sum_{\bar u \in \mathrm{orb}_{c_u}(u)} |\zeta_v^{-1}(\bar u)| \qquad (23)$$

This result allows us to relate the ATG and the sATG as follows:

Lemma 10. *$S_{QQ'}$ is a subgraph of $T_{QQ'}$*

Proof. From Equ. (23) we conclude immediately that $\eta(v, u) = 0$ implies $|\zeta_v^{-1}(\bar u)| = 0$ for all $\bar u \in \mathrm{orb}_{c_u}(u)$. Conversely, if $|\zeta_v^{-1}(\bar u)| > 0$ for at least one $\bar u \in \mathrm{orb}_{c_u}(u)$ we have $\eta(v, u) > 0$. Thus $h(v, u) = 0$ implies $h^*(v, u) = 0$ and $h^*(v, u) > 0$ implies $h(u, v) > 0$. Using Lemma 5 and Definition 10 we observe that $(u, v) \in E(S_{QQ'})$ implies $(u, v) \in E(T_{QQ'})$. ∎

Clearly, $S_{QQ'} = T_{QQ'}$ if both Q and Q' have trivial automorphism groups. This is the case if and only if all stoichiometric coefficients in r are unity and none of the involved molecules has any internal symmetries. This condition is not necessary, however:

Lemma 11. *Let $r = (Q \to Q')$ be a reaction with ATG and sATG $T_{QQ'}$ and $S_{QQ'}$, respectively. Then $T_{QQ'} = S_{QQ'}$ if and only if for each $x \in V(Q)$ there exists $y \in V(Q')$ such that $\varphi(\mathrm{orb}_Q(x)) \subseteq \mathrm{orb}_{Q'}(y)$.*

Proof. Since $V(S_{QQ'}) = V(T_{QQ'})$ by definition we only have to take care of the edge sets. Assuming $T_{QQ'} = S_{QQ'}$ yields $E(T_{QQ'}) = E(S_{QQ'})$. For an arbitrary $x \in V(Q)$ there exists exactly one $y \in V(Q')$ with $y = \varphi(x)$ due to the bijectivity of the AAM φ. With $E(T_{QQ'}) = E(S_{QQ'})$ we obtain $(\zeta^{-1}(x'), \zeta^{-1}(y')) \in S_{QQ'}$ for all $x' \in \mathrm{orb}_Q(x)$ and $y' \in \mathrm{orb}_{Q'}(y)$. This, however, implies by the definition of $S_{QQ'}$ that for each $x' \in \mathrm{orb}_Q(x)$ there exists a $y' \in \mathrm{orb}_{Q'}(y)$ such that $y' = \varphi(x')$ which yields $\varphi(\mathrm{orb}_Q(x)) \subseteq \mathrm{orb}_{Q'}(y)$. Conversely, for an arbitrary $x \in V(Q)$ there exists a $y \in V(Q')$ such that the image of the orbit of x under the bijection φ is a subset of the orbit of y. By definition of $S_{QQ'}$ this directly implies for each $x' \in \mathrm{orb}_Q(x') : (\zeta^{-1}(x), \zeta^{-1}(y')) \in E(S_{QQ'})$ for all $y' \in \mathrm{orb}_{Q'}(y)$. With x arbitrary chosen, we obtain $E(T_{QQ'}) = E(S_{QQ'})$. ■

4 Atom transition networks

The (simplified) ATGs of individual reactions can be combined to a single large *Atom Transition Network* for an entire chemical network. The construction is exactly the same for ATGs with weight $h(\,.\,,\,.\,)$ and simplified ATGs with weight $h^*(\,.\,,\,.\,)$. For the ease of presentation we will therefore simply refer to the ATG, even though the practical implementation is based on simplified ATGs for the individual reactions.

Explicit catalysts in reaction $r = (Q \to Q')$ appear as connected components both on the reactant side and the product side, i.e., c is a catalyst if $c \subseteq Q_\circ$ and there is $c' \in Q'_\circ$ such that c and c' are isomorphic. In the final ATN, c and c' cannot appear separately, thus we identify them along a label-preserving isomorphism $\mu : V(c') \to V(c)$ that maps the product to the reactant. Correspondingly, edges (u, v) with $u \in V(c)$ and $v \in V(c')$ are replaced by directed edges of the form $(u, \mu(v))$ within c. Note that this may result in loops (u, u) if there are edges (u, v) with $\mu(v) = u$. Two different reactions r_1 and r_2, moreover, may involve the same reactant c and the same product c', albeit connected with AAM that yield different

edge sets in the ATGs. Below, we therefore define the atom transition network as (weighted) multigraphs with loops.

It is also possible to model different compartments. In this case, all atom labels identify also the compartment. Exchange of material between compartments then is modeled by additional pseudo-reaction between isomorphic copies of a compound c that differ only in compartment information on their atoms. The AAM for such a transport reaction then is a graph isomorphism that preserves only atom types but not the compartment information.

Definition 11. Let $(\tilde{C}, \tilde{R})$ a be reaction network augmented by feed and drain reactions and an AAM for each reaction $r \in \tilde{R}$. The atom transition network (ATN) $A(\tilde{C}, \tilde{R})$ is the weighted multi-graph with loops that has the set $V(A) = \bigcup_{c \in C} V(c)$ of atoms in the pairwise distinct molecular graphs $c \in C$ as its vertex set – the seed and drain atoms are considered distinct from the atoms in the inner compounds $c \in C$. The edge set $E(A)$ is the disjoint union of the edges of ATG for all reactions $r = (Q \to Q') \in R$ (after merging catalysts). Each edge $(u, v)_r \in E(A)$ receives the weight $h_r(v, u)$.

By construction of the ATG for individual reactions there are edges only between atoms of the same type, and thus each weakly connected component of an ATN contains only atoms of a single type.

Lemma 12. *Let $(\tilde{C}, \tilde{R})$ be a conservative reaction network with feed and drain reactions and assume that every $c \in C \setminus C_D$ is a reactant of at least one reaction and every $c \in C \setminus C_F$ is a product of at least one reaction. Then the vertex sets of the weakly connected components of the ATN $A(\tilde{C}, \tilde{R})$ and the simplified ATN $A_S(\tilde{C}, \tilde{R})$ coincide.*

Proof. First we observe that for every reaction $r = (Q \to Q')$ and every atom $v \in V(c')$ with $c' \subseteq Q_\circ$, all $v' \in \mathrm{orb}_{c'}(v)$ are by construction in the same connected component of both the ATN and the simplified ATN. If c is the product of some reaction, then for each orbit Y, all $v \in Y$ are in the same connected component of both the ATN and the sATN. Hence this is in particular true for all internal species as well as the drain species.

Since feed reactions by construction have an isomorphism as their AAM, the statement extends to all $c \in \tilde{C}$. It therefore suffices to consider the quotient graphs Γ and Γ_s of the ATN and sATN, respectively, obtained by contracting the orbits of all compounds. By construction, Γ and Γ_S have the same vertex sets since the vertex sets and definitions of the orbits of the compounds are the same for $A(\tilde{C}, \tilde{R})$ and $A_S(\tilde{C}, \tilde{R})$. Moreover, Equ.(23) implies immediately that there is an edge (u', v') from some $u' \in \mathrm{orb}_{c_u}(u)$ to $v' \in \mathrm{orb}_{c'_v}(v')$ in $A_S(\tilde{C}, \tilde{R})$ if and only if there is an edge (u, v) in $A(\tilde{C}, \tilde{R})$. That is, Γ and Γ_S have the same edge set and thus the weakly connected components of Γ and Γ_S are the same. Making use of the observation that orbits are always contained in a single weakly connected component completes the proof. ■

Fluxes combine reactions in a manner that keeps track of the relative intensity. Given a flux F, the values of $\lambda'(v)$ of a valuation of the atoms of compound c will therefore depend on a (linear) combination of the influxes into c.

Definition 12. Let $(\tilde{C}, \tilde{R})$ a be reaction network with an AAM for each reaction $r \in \tilde{R}$, and let F be a flux on $(\tilde{C}, \tilde{R})$. Then the ATN associated with F is the ATN $A(\tilde{C}, \mathit{F}) := A(\tilde{C}, R_{\mathit{F}})$ where $R_{\mathit{F}} = \{r \in \tilde{R} | \mathit{F}(r) > 0\}$ is the set of reactions with positive flux. Each edge $(u, v) \in E(A(\tilde{C}, \mathit{F}))$ is associated with an atom-wise flux

$$\mathbf{F}_{vu} := \sum_r s^+_{c_v r} h_r(v, u) \mathit{F}(r) \tag{24}$$

Given a stationary flux F, each inner species in the network is either not involved in an active reaction at all or it appears both as reactant and product. The restriction of the CRN to the active reactions therefore always satisfies the conditions of Lemma 12, and we obtain:

Corollary. *Let F be a stationary flux on the reaction network $(\tilde{C}, \tilde{R})$. Then the vertex sets of the weakly connected components of the ATN $A(\tilde{C}, \mathit{F})$ and the simplified ATN $A_S(\tilde{C}, \mathit{F})$ coincide.*

If F is a stationary flux that is not a circulation, then $A(\tilde{C}, \mathit{F})$ has at least one feed and one drain reaction, see Section 2.2. Moreover, these

feed and drain species are uniquely identified by the flux F; assuming that a compound c is not both a feed and drain species at the same time, any violation of the flux balance condition at c uniquely identifies species with incident feed or drain reactions. Finally, each atom in a feed or drain species is attached to feed or drain atoms.

A key observation is that for steady-state fluxes F of the metabolites, flux balance also holds atom-wise on the ATN:

Lemma 13. *Let $A(\tilde{C}, F)$ be an ATN and let F be a steady-state flux. Then for all $w \in V(A(\tilde{C}, F)) \setminus (C_F \cup C_D)$ holds*

$$\sum_{u \in V(A(\tilde{C},F))} \mathbf{F}_{wu} = \sum_{v \in V(A(\tilde{C},F))} \mathbf{F}_{vw}$$

Proof. The definition of $\mathbf{F}_{uv}$ and Lemma 7 yields

$$\sum_{u \in V(A(\tilde{C},F))} \mathbf{F}_{wu} = \sum_r s^+_{c_w r} \underbrace{\sum_{c \in \tilde{C}} \sum_{u \in V(c)} h_r(w,u)}_{=1} F(r) = \sum_r s^+_{c_w r} F(r)$$

$$\sum_{v \in V(A(\tilde{C},F))} \mathbf{F}_{vw} = \sum_r \underbrace{\sum_{c \in \tilde{C}} \sum_{v \in V(c)} s^+_{c_v r} h_r(v,w)}_{=s^-_{c_w r}} F(r) = \sum_r s^-_{c_w r} F(r)$$

Since every inner compound $c \in C$ satisfies $\sum_r s^-_{cr} F(r) = \sum_r s^+_{cr} F(r)$ for every steady-state flow, the two sums coincide. ∎

Since $A(\tilde{C}, F)$ is an edge-weighted graph, it has well-defined weakly connected components. Denote by W the subset of vertices in a weakly connected component of the ATN $A(\tilde{C}, F)$. We first note that Lemma 8 remains valid with $\Lambda(c) := \sum_{u \in V(c) \cap W}$ since for any reaction with a non-zero flux we have $h_r(v,u) = 0$ unless u and v are connected by an edge in the full ATN. Finally, we note that the weakly connected components of ATNs and sATNs coincide, see Lemma 12.

5 Propagation of valuations on atom transition networks

Recall that the dynamics of concentrations of the network is given as $\dot{x} = \mathbf{S}\boldsymbol{\mathit{F}}(x)$. We can interpret this also at the level of the individual atoms, i.e., the concentration of an atom v in c is simply given by x_c. The atom-level fluxes are given by $\mathbf{F}_{vu}$, and hence the outflux from u is therefore $D_u^{out} \coloneqq \sum_{v \neq u} \mathbf{F}_{vu}$. The corresponding Laplacian $\mathbf{L}$, thus, has entries

$$\mathbf{L}_{vu} \coloneqq D_u^{out}\delta_{vu} - \mathbf{F}_{vu}\,. \tag{25}$$

To simplify the presentation below we link all atoms $v \in V(c)$ of the external drain species $c \in C_D$ to an additional formal drain node $\varnothing$ that serves as unique absorbing state. In particular, $D_{\tilde{u}}^{out}$ for an atom $\tilde{u} \in V(\tilde{c})$ with $\tilde{c} \in C_D$ is the outflow into the drain $\varnothing$ and thus equals the inflow $\mathbf{F}_{\tilde{u}u}$ through transport reaction $(c \to \tilde{c})$.

In order to model the time evolution of the valuations $\lambda(u)$ we assume that λ is propagated proportional to the flux $\boldsymbol{\mathit{F}}(r)$ through every reaction r and consider the time evolution of $\gamma_u \coloneqq \lambda(u)x_c$. In the general case we therefore have $\dot{\gamma}_u = x_c\dot{\lambda}(u) + \dot{x}_c\lambda(u)$. We are in particular interested in the case that $\boldsymbol{\mathit{F}}$ is a steady-state flux, while the isotope labeling modeled by the valuation λ is newly introduced into the system. In this case x_c is the constant steady-state concentration of compound c.

We assume that the feed species are replenished by means of an external flux which in turn gives rise to an inflow $\varpi(v)$ of the valuation $\lambda(v)$ to the atoms v of compounds $c_v \in C_F$. Under these assumptions, the time evolution of λ has the form

$$x_{c_v}\dot{\lambda}(v) = \sum_{u \in V(A(\tilde{C},\boldsymbol{\mathit{F}}))} (-\mathbf{L}_{vu})\lambda(u) + \varpi(v) \tag{26}$$

for all $v \in V(A(\tilde{C},\boldsymbol{\mathit{F}}))$. Moreover, $\varpi(v) = 0$ unless $v \in V(c_v)$ with $c_v \in C_F$, i.e., v is an atom of a feed molecule. Note that the formal drain $\varnothing$, i.e., the absorbing state, does not appear in this description.

In the following it will be convenient to consider the weakly connected

components of $A(\tilde{C}, F)$ separately. In particular, we will disregard components that are not connected to food and drain atom.

Proposition 2. *Let $A(\tilde{C}, F)$ be a weakly connected ATN with feed and drain atoms. Then the matrix $-\mathbf{L}$ is Hurwitz stable, i.e., the real parts of all its eigenvalues are negative. Moreover, $\mathbf{L}^{-1}$ has only non-negative entries.*

Proof. The matrix $\mathbf{L}$ can be rearranged such that each strongly connected component B of $A(\tilde{C}, F)$ appears consecutively. Denote the restriction of $\mathbf{L}$ to B by $\mathbf{L}_B$. Since F is not a circulation, every atom in $A(\tilde{C}, F)$ is connected by a path to $\varnothing$ and thus every strongly connected component B contains an atom w that drains either to a different block or, the compound c_w harboring w is a drain species, $c_w \in C_D$, then w drains to $\varnothing$. Thus $\mathbf{L}$ cannot be written in block-diagonal form.

By construction $\mathbf{L}$ is a weakly diagonally dominant, i.e., $D_w^{out} \geq \sum_v \mathbf{F}_{vw}$. The inequality is strict exactly for the feed atoms $w \in c_w$ with $c_w \in C_F$. For each irreducible block $\mathbf{L}_B$, furthermore, there is at least one species w with $D_w^{out} > \sum_{v \in B} \mathbf{F}_{vw}$. Thus $\mathbf{L}_B$ is *irreducibly diagonally dominant* for each B. Moreover the diagonal entries $\mathbf{L}_{uu} = D_u^{out}$ are strictly positive, i.e., $\mathbf{L}$ is a so-called L-matrix. It is well known that all eigenvalues of an irreducibly diagonally dominant L-matrix have positive real parts [59]. Moreover, the block-triangular structure of $\mathbf{L}$ implies that its spectrum is the union of the spectra of the blocks $\mathbf{L}_B$. Thus all eigenvalues of $\mathbf{L}$ have a positive real part.

Using e.g. Thm.2.2.4 of [7], a non-singular weakly diagonally dominant L-matrix is equivalent to a weakly diagonally dominant non-singular M-matrix. Alternatively, it is not difficult to verify that $\mathbf{L}$ is a "weakly chained diagonally dominant" L-matrix, or "a matrix of positive type" [12], which implies that it is non-singular [8, 56] and an M-matrix [7]. Therefore $\mathbf{L}$ satisfies the fifty properties listed in Thm.(2.3) of [10]. In particular, $\mathbf{L}$ is "inverse positive", i.e., $\mathbf{L}^{-1}$ has only non-negative entries. ■

Let us write $\mathbf{X}$ for the diagonal matrix of atom-wise concentrations, i.e., $X_{vw} \coloneqq x_{c_w} \delta_{vw}$ for all $v, w \in V(A(\tilde{C}, F))$. Considering λ as a column vector indexed by the atoms $v \in V(A(\tilde{C}, F))$, we may rewrite Equ.(26) in

the form

$$\dot{\lambda} = \mathbf{A}\lambda + b \tag{27}$$

with $\mathbf{A} \coloneqq -\mathbf{X}^{-1}\mathbf{L}$ and $b_v \coloneqq \varpi(v)/x_{c_v}$, i.e., $b = \mathbf{X}^{-1}\varpi$.

The solutions of ordinary differential equations of the form Equ.(27) are well known [62].

Proposition 3. *If* $\mathbf{A}$ *is Hurwitz stable and* b *is a constant, then the initial value problem Equ. (27) with initial condition* $\lambda(t = 0) = \lambda_0$ *is solved uniquely by*

$$\lambda(t) = \lambda_\infty + \exp(t\mathbf{A})(\lambda_0 - \lambda_\infty) \qquad \textit{with} \quad \lambda_\infty = -\mathbf{A}^{-1}b\,. \tag{28}$$

Moreover, the stationary solution λ_∞ *is globally stable.*

Proof. In brief, the coordinate transformation $\mu \coloneqq \lambda + \mathbf{A}^{-1}b$ reduces Equ.(27) to the linear ODE $\dot{\mu} = \mathbf{A}\mu$. ■

It remains to show that $\mathbf{A} = -\mathbf{X}^{-1}\mathbf{L}$ is also stable. Using well known results we may state:

Proposition 4. *Let* $\mathbf{M}$ *be a matrix with non-positive off-diagonal entries. If the real parts of all eigenvalues of* $\mathbf{M}$ *are positive then the real parts of all eigenvalues of* $\mathbf{DM}$ *are positive for positive diagonal matrices* $\mathbf{D}$.

Proof. Since the real parts of all eigenvalues are positive, $\mathbf{M}$ is indeed a non-singular M-matrix by Thm.(2.3) in [10] and hence in particular all principal minors are positive. As noted e.g. in [37], this property as well as the non-negativity of the off-diagonal entries are preserved under multiplication with $\mathbf{D}$, and thus using Thm.(2.3) in [10] again, $\mathbf{DM}$ is also a non-singular M-matrix and hence the real parts of all its eigenvalues are again positive. Alternatively, [9, 30] showed that stability and D-stability are equivalent for Metzler matrices, i.e., matrices with non-negative off-diagonal elements. ■

As a consequence of Prop. 2 the Laplacian $\mathbf{L}$ satisfies the conditions of Prop. 4, and hence we obtain:

Corollary. *Let* $\mathbf{L}$ *be the Laplacian of a weakly connected ATN with source and drain atom. Then the matrix* $\mathbf{A} \coloneqq -\mathbf{X}^{-1}\mathbf{L}$ *is Hurwitz stable,* $\mathbf{A}^{-1}$ *has only non-positive entries, and hence* λ_∞ *is non-negative for all non-negative vectors* b*.*

A practical problem with applying this result directly is that it involves the steady-state concentrations x_c of all metabolites c. It is well known, however, that these cannot be computed from the structure of the ATN and the steady-state fluxes alone. Additional information, most prominently Gibbs energies for the reactions, is required in general, see e.g. [60]. We will therefore avoid the assumption of concrete values of x_c as much as possible in the following.

Hence start from $0 = -\mathbf{L}\lambda + \varpi$. We first consider $u \in V(c_u)$ with $c_u \in C_F$. By assumption, we only have a single outflux reaction from c_u, namely the transport from u to its counterpart u' in the same molecule at the inside of the system with flux $F(c_u \to c'_u)$. Thus we have $D_u^{out} = \mathbf{F}_{u'u}$. The only influx is the replenishing from $\varnothing$, which does not appear as an atom in the ATN $A(\tilde{C}, F)$. As implicitly defined in Equ.(26), $\varpi(u)$ subsumes the valuation transported to the feed atom u. By flux balance, every steady-state solution satisfies $D_u^{out}\lambda_\infty(u) = \varpi(u)$. Thus it suffices to solve $[\mathbf{L}\lambda_\infty]_v = 0$ for the non-feed atoms v with the values $\lambda_\infty(u)$ already fixed for the feed atoms. In this manner we completely avoid using $\mathbf{X}$.

Before we proceed, we note some important consequences of the fact that $\mathbf{A}^{-1}$ has only non-positive entries. Since the influx ϖ and thus also b is non-negative and has positive entries only for feed atoms, we can write the steady-state flow λ_∞ as

$$\begin{aligned} \lambda_\infty(u) &= \sum_{w \in V(C_F)} \lambda_\infty(w)\vartheta_w(u) \qquad \text{with} \\ \vartheta_w &\coloneqq -\mathbf{A}^{-1}b_w = \mathbf{L}^{-1}\mathbf{X}\mathbf{X}^{-1}\varpi_w = \mathbf{L}^{-1}\varpi_w \end{aligned} \tag{29}$$

where b_w is the inflow into the single atom $w \in V(C_F)$ adjusted such that $\lambda_\infty(w) = 1$, i.e., $\varpi_w(w) \coloneqq D_w$ and $\varpi_w(w') \coloneqq 0$ for $w' \neq w$. We may think of ϑ_w as a basis labeling pattern obtained from marking a single influx atom. Inverse positivity of $\mathbf{L}$ ensures that ϑ_w is a non-negative vector.

As an immediate consequence we note that the steady-state solutions are monotonic in the following sense:

Corollary. *Let λ_∞ and λ'_∞ be two stationary solutions of Equ.(26) for a weakly connected ATN with source and drain atom and assume $\lambda_\infty(w) \geq \lambda'_\infty(w)$ for all seed atoms $w \in V(C_F)$. Then $\lambda_\infty(u) \geq \lambda'_\infty(u)$ for all $u \in V(A(\tilde{C}, F))$.*

Increasing the labeling fraction $\lambda_\infty(w)$ of some seed atom $w \in V(C_F)$ therefore never decreases the stationary labeling fraction $\lambda_\infty(u)$ of any other atom u in the ATN.

It remains to consider the special case that all feed atoms are fully labeled, i.e., that $\lambda_\infty(w) = 1$ for all $w \in C_F$. The following result shows that in this case the entire part of the ATN that is reachable from the feed is flooded with the marks in the feed.

Theorem 5. *Consider λ_∞ being a stationary solution of Equ.(26) for a weakly connected ATN with source and drain atom and let $\lambda_\infty(w) = 1$ for all feed atoms $w \in V(C_F)$. Then $\lambda_\infty(u) = 1$ for all $u \in V(A(\tilde{C}, F))$.*

Proof. It suffices to show that the all-one vector is a solution of $\mathbf{L}\lambda_\infty = \varpi_1$, where $\varpi_1(w) = D_w^{out}$ for $w \in V(C_F)$ and $\varpi_1(w) = 0$ otherwise, because uniqueness is ensured by Prop.(3) and the influx vector ϖ_1 is adjusted such that $\lambda_\infty(w) = 1$ for all $w \in V(C_F)$ as required. Writing $\vec{1}$ for the all-one column vector, and $V \coloneqq V(A(\tilde{C}, F))$ we have $[\mathbf{L}\vec{1}]_w = \sum_{u \in V} \mathbf{L}_{wu} = D_{ww} = \varpi(w)$ for $w \in V(C_F)$ since feed atoms have no other influx than $\varpi(w)$. For $v \in V \setminus (V(C_F) \cup V(C_D))$ flux balance implies

$$[\mathbf{L}\vec{1}]_v = \sum_{u \in V} \mathbf{L}_{vu} = D_{vv} - \sum_{u \neq v} \mathbf{F}_{vu} = \sum_{u \neq v} \mathbf{F}_{uv} - \sum_{u \neq v} \mathbf{F}_{vu} = \varpi(v)\,.$$

For the drain vertices $w \in C_D$, finally, we have $D_{ww} = \sum_{u \neq w} F_{wu}$ by assumption, and thus $[\mathbf{L}\vec{1}]_w = 0$. Thus $\vec{1}$ is indeed a solution of the steady equation $\lambda_\infty \mathbf{L} = \varpi_1$, and hence $\lambda_\infty = \vec{1}$. ∎

Since $\mathbf{A}$ is monotone (i.e., $\mathbf{A}^{-1}y$ is non-positive for any vector y with non-positive entries) and Hurwitz stable, $\exp(t\mathbf{A})y$ is coordinate-wise decreasing for any non-negative vector y. Thus $\lambda(t)$ remains between 0 and

1 provided this is true for λ_0 and λ_∞. Since the external flow ϖ_w is used here only as a formal device to model a temporally constant labeling of the feed atoms, we assume $\lambda_0(w) = \lambda_\infty(w)$ for all $w \in V(C_F)$.

A valuation λ, thus, can be interpreted as the (expected) fraction of a label, provided $\lambda_\infty(w) = \lambda_0(w) \in [0, 1]$ for the atoms $w \in V(C_F)$ of the feed set. The dynamics of Equ. (28) then describes how a steady-state flux acts to distribute the label throughout the weakly connected component of the ATN that contains w. Usually, we will assume $\lambda_0(u) = 0$ for the non-feed atoms $u \notin V(C_F)$. This assumption is not necessary, however, and may be replaced for instance by the natural isotope distribution.

We finally show that the graph-theoretic structure of the ATN alone determines the set of atoms that are eventually reached by isotope labels. To this end we consider the stationary labeling patterns ϑ_w defined in Equ. (29) for single labeled feed atom:

Theorem 6. *Let $A(\tilde{C}, \tilde{R})$ be an ATN with feed set C_F and a feed atom $w \in V(C_F)$. Then for all $u \in V(A(\tilde{C}, \tilde{R}))$ holds $\vartheta_w(u) > 0$ if and only if u is reachable from w.*

Proof. Denote by T the breadth-first search tree in $A(\tilde{C}, \tilde{R})$ rooted at $w \in C_F$. By construction we have $(\vartheta_w(w) = 1)$. Moreover, we know that $\vartheta_w(u) \geq 0$. We proceed by induction and assume that at every step of BFS traversal $\vartheta_w(u) > 0$ holds for all vertices visited so far. Let v be an unvisited out-neighbor of u. Then there is a reaction $r = (Q \to Q')$ with positive flow F_r such that $h_r(u, v) > 0$ and hence in particular $\mathbf{F}_{vu} > 0$. In particular, $v \notin V(C_F)$ since feed atoms have no incoming arcs (except from the formal source $\varnothing$, which we are not considering). The stationary distribution satisfied $[\mathbf{L}\vartheta_w]_v = 0$, and thus

$$\vartheta_w(v) D_v^{out} = \sum_{u'} \mathbf{F}_{vu'} \vartheta_w(u') \geq \mathbf{F}_{vu} \vartheta_w(u) > 0$$

where we use the induction hypothesis $\vartheta_w(u) > 0$. Since $D_v^{out} > 0$ for all v, we have $\vartheta_w(u) > 0$ for all u reachable from w. For the converse we first note that we can restrict ourselves to the weakly connected component of the ATN that contains the feed for w. If there is more than

one component, $\mathbf{L}$ can be written in block-diagonal form, which is also true for $\mathbf{L}^{-1}$. Since only the block containing w has an influx ϖ, we obtain $\vartheta_w(u) = [{\mathbf{L_k}}^{-1}\vec{o}]_u = 0$ for all u in all other weakly connected components. The vertex set W of the component containing w is naturally partitioned into a set W_1 reachable from w and its complement W_2. By assumption, there is no edge from W_1 into W_2 and thus the Laplacian has block-triangular form and all vertices u with $\varpi(u) > 0$ belong to W_1. The equation for the steady-state flow on W_2 thus becomes $\mathbf{L}(W_2)\lambda_\infty = 0$, and the restriction of ϑ_w to W_2 vanishes since $\mathbf{L}\vartheta_w = b$ has a unique solution. ■

That is, reachability in the ATN, and thus the graph structure of the ATN alone, already specifies all atoms that will eventually carry an isotope label.

6 Superpositions of pathways

A straightforward application of isotope labeling is to distentangle alternative pathways. At least four well-characterized pathways have been described in the literature that collectively amount to the conversion of glucose into pyruvate: In addition to the classic model of glycolysis, the Embden-Meyerhoff-Parnas (EMP) pathway [52], the Entner-Doudoroff (ED) pathway [18], the oxidative Pentose Phosphate Pathway (PPP) [61], as well as the generation of glycerinaldehyd 3-phosphate and subsequent conversion using Transketolase-1 (TKT-1) are wide-spread [14, 18, 57]. The four pathways yield distinct labeling patterns in the end product pyruvate depending on the position of the label in the glucose feed, see Figure 5 and also [42].

During the oxidative part of the PPP, 6-phosphogluconate is decarboxylated to ribulose 5-phosphate, thereby loosing the isotopic labeled carbon atom if $^{13}C_1$-glucose is engaged as isotopic tracer. Hence, pyruvate generated purely via the PPP (oxidative part as well as TKT-1) is completely unlabeled. In contrast, ED and EMP each yield 50% single-labeled and 50 % unlabeled pyruvate. The labels, however, emerge at different po-

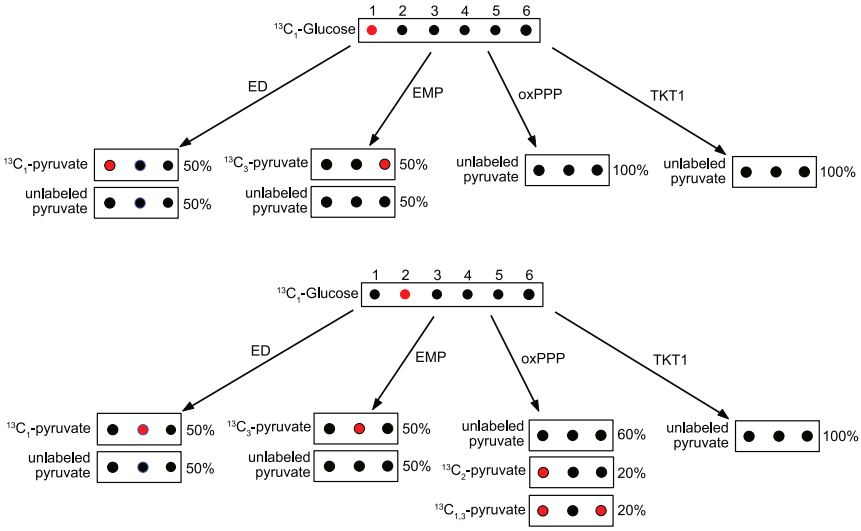

Figure 5. Schematic depiction of the degradation of single labeled $^{13}C_1$-glucose and $^{13}C_2$-glucose via four different metabolic pathways: Entner-Doudoroff (ED), Entner-Meyerhoff (EMP), oxidative Pentose Phosphate Pathway (PPP), and generation of glycerinaldehyd 3-phosphate as well as subsequent susbtrates by Transketolase-1 (TKT-1).

sitions due to different catalytic mechanisms and intermediates. ED and EMP, therefore, are indistinguishable via conventional, MID based, mass spectrometry but exhibit distinct positional enrichment: The carbon atom in pyruvate with the highest oxidation state exhibits positional enrichment of 50% when $^{13}C_1$-glucose is degraded via ED and all the others carbon are fully depleted. In contrast, degradation of $^{13}C_1$-glucose via EMP yields a positional enrichment of 50% for the third carbon atom only. $^{13}C_2$-glucose, on the other hand, is not suitable to discriminate between the two pathways even in positional enrichment analysis. Isotopomers generated via TKT-1 are unlabeled, while the oxidative part of the PPP yields pyruvate isotopomers single-labeled at position 1 or double-labeled at position 1 and 3, additionally. Nevertheless, discrimination of glycolysis from PPP is of great importance since it reflects the ability of cells to cope with oxidative stress. Tumor cells, for example, exhibit higher potential of PPP usage in order to generate NADPH for glutathione reduction and elimination of

oxidative species such as H_2O_2 [58].

In many organisms more than one glycolytic pathway may be active concurrently, see e.g. [47]. It is therefore of interest to disentangle the relative contributions of distinct pathways. To this end, we first note that each pathway can be thought of as a flux $F^{(k)}$ normalized such that the influx from a single feed species $\varnothing_c \in C_F$ is set to unity, i.e., $F^{(k)}_{\varnothing_c \to c} = 1$. In this case $D_u^{out(k)} = 1$ for each atom $u \in V(\varnothing_c)$. It is important to keep in mind that such a normalized pathway is a purely structural feature of the underlying chemical reaction network. The input labeling ϖ is used by all competing pathways indiscriminately.

Pathway Deconvolution Problem. *Given a set of m normalized pathways $F^{(k)}$, $k = 1, 2, \ldots, m$, and a known input labeling $\varpi_\bullet$ on the feed compounds C_F, can we identify the relative fluxes through the m competing pathways from measurements of the steady-state labeling pattern λ_∞?*

Let us denote by $\hat{\mathbf{L}}^{(k)}$ the Laplacian of the normalized pathway $F^{(k)}$ and let $\varpi^{(k)}$ denote the corresponding normalized input flux. Additivity of flows allows us to consider the weighted superposition of the pathways:

$$\mathbf{L} \coloneqq \sum_{k=1}^{m} \alpha_k \hat{\mathbf{L}}^{(k)} \quad \text{and} \quad \varpi_\bullet \coloneqq \sum_{k=1}^{m} \alpha_k \varpi^{(k)} \quad \text{with} \quad \alpha_k \geq 0 \text{ for } 1 \leq k \leq m\,. \tag{30}$$

Since the absolute value of the flux only affects the time-scale in Equ.(26) but not the steady-state vector λ_∞ we may require, in addition, that $\sum_{k=1}^{m} \alpha_k = 1$. Then

$$\lambda_\infty = \left(\sum_{k=1}^{m} \alpha_k \hat{\mathbf{L}}_{(k)} \right)^{-1} \varpi_\bullet \tag{31}$$

Importantly, it is **not** a linear combination of labeling patterns generated by the individual pathways, which would be computed as $\lambda_\infty^{(k)} = \mathbf{L}_{(k)}^{-1} \varpi_k$. The reason for the non-linear dependence is that labels are mixed at each intermediate at which pathways intersect.

Measurements of $\lambda_\infty(v)$ for a sufficient number of well-chosen atoms of

intermediates or drain species can be used to determine the relative flows α_k through the different pathways.

Proposition 7. *The pathway deconvolution problem has either a unique solution or a linear manifold of degenerate solutions.*

Proof. Using the fact that the matrix $\mathbf{L}$ is invertible we can rewrite Equ. 31 in the form

$$\sum_{k=1}^{m} \alpha_k \left[\hat{\mathbf{L}}_{(k)} \lambda_\infty\right]_u = \varpi_\bullet(u)\,. \tag{32}$$

The vector $(\alpha_1, \ldots, \alpha_m)$ is therefore given by a system of linear equations that includes the additional constraint that the α_k add to unity. Standard results on systems of linear equations immediately imply the claim. ■

If there is a unique solutions, moreover, $m - 1$ well-chosen, i.e., independent measurements together with the normalization conditions are sufficient to completely determine the relative activity of the individual pathways.

In more complex examples examples, successful labeling strategies may depend on the metabolic state of the cell. As an example, we consider the synthesis of isopentenyl pyrophopshate (IPP) from pyruvate and glycerinaldehyde 3-phosphate (GA3P) [63]. There are two major synthesis pathways, the mevalonate pathway (MVA) and the MEP pathway, see Figure 6. The MVA pathway involves the generation of β-hydroxy-β-methylglutaryl-CoA (HMG-CoA), which requires two acetyl-CoA molecules for acetoacetyl-CoA and one additional acetyl-CoA is used subsequently to synthesize HMG-CoA. This multiple use of acetyl-CoA accounts for the complicated structure of the ATN in this case and the generation of multiple isotopomers depending on the particular labeling strategy.

Labeling strategies using two distinct isotopomers of pyruvate as isotopic tracers are highlighted in Fig. 6 by red and blue edges, respectively. In an anabolic metabolic state gluconeogenesis from pyruvate can be considered negligible and hence GA3P remains unlabeled. In this situation, both choices of the label discriminate the two pathways in a manner that is measurable both by positional enrichment and MID measurements. However, the situation becomes more complicated if the cells depend on glu-

cose. Deployment of $^{13}C_1$-glucose as isotopic tracer generates three different pyruvate isotopomers (see Fig.5). If only labeled pyruvate were used for IPP synthesis via the MVA pathway the resulting IPP isotopomers would coincide with the ones generated by red labeling strategy. However, three different pyruvate isotopomers give rise to two acetyl-CoA isotopomers since the label obtained at the first carbon atom in pyruvate generated via ED is lost due to decarboxylation. Thus, in total, $2^3 = 8$ different HMG-CoA isotopomers emerge, which give rise to eight different IPP isotopomers, i.e. unlabeled, $^{13}C_1$, $^{13}C_3$, $^{13}C_5$, $^{13}C_{1,3}$, $^{13}C_{1,5}$, $^{13}C_{3,5}$, and $^{13}C_{1,3,5}$-IPP. The MEP pathway, however, depends additionally on GA3P, which emerges in two flavors, unlabeled and as $^{13}C_1$-GA3P. Propagation of the additional label introduced via $^{13}C_1$-GA3P is depicted in orange. Hence, next to unlabeled IPP, three other isotopomers are generated, i.e. $^{13}C_2$, $^{13}C_5$, and $^{13}C_{2,5}$-IPP. Obviously, the diversity of isotopomers would increase even more, if $^{13}C_2$- glucose instead of $^{13}C_2$-glucose were employed as isotopic tracer (see figure 5). In such a complex situation, solving the pathway deconvolution problem aids to compute the contribution of each metabolic pathway.

However, these examples highlight the fact that the structure of the ATN even without a precise computation of λ_∞ is informative in the sense of determining the atoms that will eventually be labeled. In particular, the graph-theoretic information about the ATN is sufficient to reason about labeling strategies at a qualitative level.

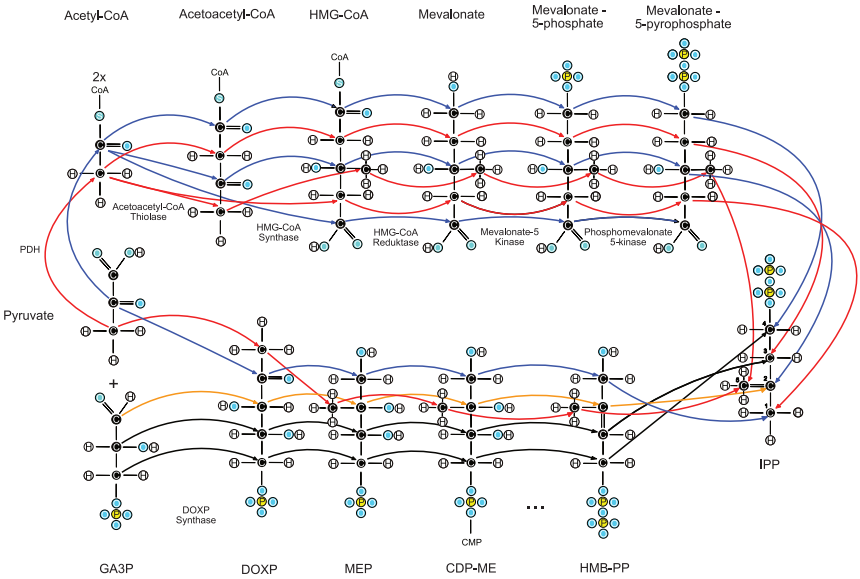

Figure 6. Trace of isotope labels distinguishing the mevalonate pathway (MVA, above) and non-mevalonate pathway (MEP, below) for isopentenylpyrophosphate (IPP) synthesis. Colored edges indicate labeling transmission for different labeling strategies: $^{13}C_2$-pyruvate (blue) and $^{13}C_3$-pyruvate (red), respectively. Transition edges depicted in orange indicate additional marks introduced by GA3P if $^{13}C_1$-glucose instead of $^{13}C_3$-pyruvate (red) is used. Depending on the pathway and labeling strategy multiple isotopomers can be found (see below). Abbreviations for enzymes and compounds: PDH: Pyruvatedehydrogenase, CoA: Coenzym A, HMG: 3-Hydroxy-3-Methylglutaryl, GA3P: glycerinadehyd 3-phosphate, DOXP: 1-deoxy-D-xylulose 5-phosphate, MEP: methyl-D-erythritol phosphate, CDP-ME: 4-diphosphocytidyl-2-C-methylerythritol, HMB-PP: E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate, IPP: Isopentenylphosphate.

7 Multi-labeling experiments

Some metabolic pathways might not be discriminable with single labeled isotopic tracers. Multi-labeling experiments can provide more precise metabolic pathway differentiation in such cases and address additional research questions [34]. The isotopic tracers deployed in such a setting can be of one or different compounds (e.g. glucose and glutamine). In particular, glucose and glutamine dependent isotopic tracers are engaged to study fluxes in initial glucose breakdown (glycolysis and PPP) and TCA, respectively. In addition, multiple ($^{13}C_{1,2}$-glucose) or all (U-[^{13}C]-glucose, U-[^{13}C]-glutamine) atoms of interest can be isotopic labeled.

While multi-labeled tracers aid to enhance the separation resolution of metabolic pathway usage [3], some pathways, generating the same intermediate compound from a common substrate via different catalytic mechanisms, might only be distinguishable with multi-labeled tracers. Here, also parallel labeling experiments might be suitable [19]. Fully labeled tracers serve to study the global metabolic activity of cells and examine which substrates are generated from the feeding substance. If in one experiment carbon as well as aminoacid metabolism is of interest, heterogenously labeled isotopic tracers such as U-[^{13}C]-U-[^{15}N]-glutamine can be of interest.

The theory outlined in the previous sections allows the computation of positional enrichment both for single-labeled and multiple labeled isotopic tracers. If more than one feed atom in the inflow is labeled, then the resulting stationary distribution λ_∞ is obtained as the sum of these contributions according to Equ. (29):

Corollary. *Let $A(\tilde{C}, \tilde{R})$ be an ATN of a conservative reaction network $(\tilde{C}, \tilde{R})$ and assume that the feed species $c \in C_F$ are labeled with temporally constant tracer fractions $\lambda_\infty(w)$ for $w \in V(c)$. Then the system asymptotically approaches the stationary labelling patterns*

$$\lambda_\infty(u) = \sum_{w \in V(C_F)} \lambda_\infty(w) \vartheta_w(u)$$

where ϑ_w is the labeling pattern obtained by exclusively labeling the feed atom w.

We observe that the asymptotic labeling λ_∞ is independent of the steady-state concentrations $\mathbf{X}$ even though the speed at which the equilibrium state is reached does depend on these, usually unknown, values.

An important caveat is that the formalism derived here correctly computes the labeling intensity $\lambda(u)$ for a single atom, but does not account for correlations between labeling intensities between two or more atoms in the same compound. In particular, therefore, λ is not sufficient to compute MIDs.

8 Discussion

We introduced here a formal description of atom transition networks as graphs weighted by transition probabilities that exactly account for symmetries of the reactant and product molecules as well as stoichiometry. Equ.(12) – or its sparsified variant Equ.(21) – can be readily evaluated given the graph representations of the reactants $c \in V(Q_\circ)$ and reaction products $c' \in V(Q'_\circ)$ together with an AAM φ for each reaction $(Q \to Q')$. The main purpose of the work reported here is to provide an explicit construction of the ATN both as a simple graphs and as weighted graphs. The key step is the definition of weighted atom transition graphs, ATG, for individual reactions. At this level, both symmetries of the individual compounds and non-trivial stoichiometry, which is also a form of symmetry, are incorporated completely. The ATN is then obtained as a quotient graph of the union of the molecule *complexes*, ensuring that the vertices in the ATN appear partitioned into equivalence classes whenever their fate cannot be separated. As an example, consider the transitions from acetoacetyl-CoA in Fig. 6. Although target atoms in acetoacetyl-CoA are not symmetric, each carbon atom of acetyl-CoA exhibits two outgoing transition edges for the reaction catalyzed by acetoacetyl-CoA thiolase.

Taking symmetries of molecules into account is essential for determining positional enrichment also in non-symmetric compounds. Figure 7 depicts the sATN with and without symmetries generating oxalacetate via the TCA from glutamine. In particular, the consequences of neglecting symmetries in molecules such as fumarate or succinate not only af-

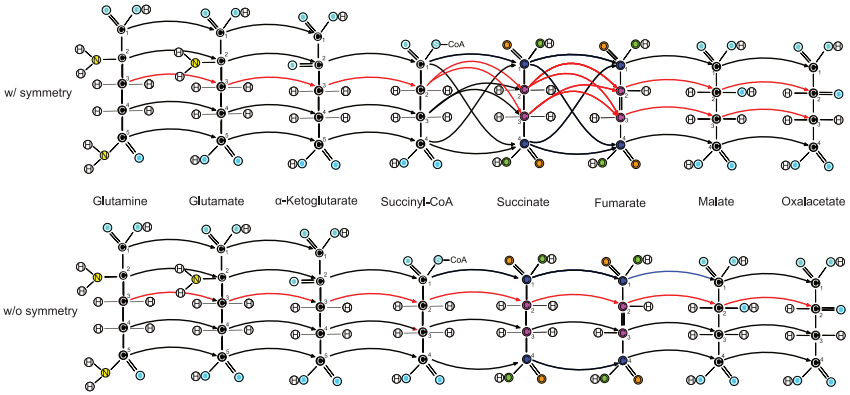

Figure 7. Illustration of the metabolism and atom-to-atom transition of glutamine to oxalacetetate with and without consideration of symmetries. Symmetric atoms are depicted in the same color. Transition edges distributing a labeling are marked red.

fects molecules harboring symmetric carbon atoms. Indeed, labeling of the third carbon atom of glutamine results in 50% labeling of the second and third carbon atom of the unsymmetric oxalacetate when symmetries are respected, although both atoms are distinguishable. In contrast, solely engaging one of the possible 16 AAMs, as depicted in the lower panel, would yield a fully enriched second carbon atom of oxalacetate while the third one would remain fully depleted. However, since carbon atoms 2 and 3 in succinate are indistinguishable, the percentage of labels on these atoms is naturally identical.

For practical applications, the formalism outlined here relies on three ingredients: (i) the efficient computation of the orbits for each compound $c \in C$, (ii) a way to compute steady-state fluxes for a particular metabolic state, and (iii) the availability or computational prediction of AAMs. The first task is readily solved by constructing a set of generators of the automorphism group $\mathrm{Aut}(c)$ for each $c \in C$; software for this purpose is readily available, including `nauty` [40] or the framework described in [2]. Stereochemistry can be accomodated without problems by replacing the automorphism group $\mathrm{Aut}(c)$ with the actual symmetry groups of the molecule.

The second task is a topic of MFA and FBA.

Reliable AAMs are the most difficult problem. Directly experimentally verified AAMs are scarce and the chemically correct AAMs cannot, in general, be obtained as the solution of combinatorial optimization problems such as the minimization of "chemical distance" between reactants and products [36] or one of the variants of maximum-common subgraph (MCS) problems [21]. The most recent generation of AAM prediction tools, which are heavily based on machine learning, provide significant advances [17, 45, 54] but still perform much less than perfectly [38, 48]. Predicted isotope labeling patterns provide a fairly direct avenue to test AAMs by comparing measurements and predictions for reactant and product labeling patterns.

We suggest that the material presented here will also serve as starting point towards addressing several related research topics. A relatively straightforward generalization is to consider time-resolved labeling experiments. A major advantage of the formalism presented here is that it can be expressed entirely in terms of linear algebra. This remains true for the case of non-stationary data. In fact, it suffices to consider the solutions of Equ.(27) for temporally non-constant in-fluxes $b(t)$. These can be written explicitly as

$$\lambda(t) = \int_0^t \exp((t-s)\mathbf{A})b(s)ds = \exp(t\mathbf{A}) \int_0^t \exp(-s\mathbf{A})b(s)ds + \exp(t\mathbf{A})\lambda_0 \tag{33}$$

and describe e.g. the dynamics of a pulse $b(t)$. Such time-dependent results are of great interest for experimentalists since research questions not only focus on isotopic equilibrium. Metabolic pathway utilization and flux-estimation might be more accurate based on non-stationary labeling analysis due to larger amount of experimental data and the absence of superimposed enrichments of labeled material from side branches. This is relevant in particular for systems which are slow to label or for quantification of substrate cycles [16, 43, 44, 64, 71]. In addition, metabolic fluxes of physiological responses to stress, such as wound healing in potato tubers, can be transient and therefore not detectable via steady-state labeling [51]. Some pathways might also be inaccessible at the beginning of a labeling

process. Another reason to rely on non-stationary data is that, as proven in Thm. 5, deployment of fully labeled isotopic tracers would yield the same information as unlabeled substrates in isotope stationarity. Indeed, for organisms depending on monoisotope-tracers, such as plants on CO_2 [28, 55] or *Pichia pastoris* on CH_3OH [15], stationary isotope solutions do not provide any information on fluxes.

Positional enrichment refers to a single atom. As noted above, it is not sufficient to correctly determine the MID since λ cannot account for correlations between labels co-occurring in the same compound. Different extensions have been introduced to deal with this problem without resorting to complete explicit enumeration of all isotopomers. The key idea is to identify fragments of metabolites that are transmitted from influx to outflux as units through all active pathways, see [50] for a formal treatment. The most widely used approaches being *elementary metabolite units* (EMU) [6] and *cumomers* [65]. In either case, positional enrichment corresponds to the lowest order data. The cumomer transition networks (CTN) for size-one cumomers, in particular, corresponds to the ATN investigated here. This suggests to consider generalizations, i.e. the transmission of subsets of atoms. In the absence of symmetries, the combination of $h(v, u)$ and $h(v', u')$ suffices to describe the transmission of pairs of atoms. Symmetries, however, introduce complications that go beyond the scope of this contribution.

***Acknowledgment*:** This work was support in part by the Novo Nordisk Foundation (grant no. 0066551, MATOMIC), the German Research Foundation as part of the SFB1052 (grant no. 209933838), and the German Federal Ministry of Education and Research (BMBF) within the German Network for Bioinformatics Infrastructure (de.NBI) under grant number w-de.NBI\303\203018.

References

[1] J. L. Andersen, C. Flamm, D. Merkle, P. F. Stadler, Chemical transformation motifs – modelling pathways as integer hyperflows, *IEEE/ACM Trans. Comp. Biol.* **16** (2019) 510–523.

[2] J. L. Andersen, D. Merkle, A generic framework for engineering graph canonization algorithms, *J. Exp. Algor.* **25** (2020) 1–26.

[3] M. R. Antoniewicz, 13C metabolic flux analysis: Optimal design of isotopic labeling experiments, *Curr. Opinion Biotech.* **24** (2013) 1116–1121.

[4] M. R. Antoniewicz, Dynamic metabolic flux analysis—tools for probing transient states of metabolic networks, *Curr. Opinion Biotech.* **24** (2013) 973–978.

[5] M. R. Antoniewicz, A guide to metabolic flux analysis in metabolic engineering: Methods, tools and applications, *Metab. Engin.* **63** (2021) 2–12.

[6] M. R. Antoniewicz, J. K. Kelleher, G. Stephanopoulos, Elementary metabolite units (EMU): A novel framework for modeling isotopic distributions, *Metab. Engin.* **9** (2007) 68–86.

[7] P. Azimzadeh, A fast and stable test to check if a weakly diagonally dominant matrix is a nonsingular M-matrix, *Math. Comput.* **88** (2019) 783–800.

[8] P. Azimzadeh, P. A. Forsyth, Weakly chained matrices, policy iteration, and impulse control, *SIAM J. Num. Anal.* **54** (2016) 1341–1364.

[9] A. Berman, D. Hershkowitz, Matrix diagonal stability and its implications, *SIAM J. Alg. Discr. Meth.* **4** (1983) 377–382.

[10] A. Berman, R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Soc. Industrial Appl. Math., Philadelphia, PA, 1994.

[11] C. L. Boulangé, A. L. Neves, J. Chilloux, J. K. Nicholson, M.-E. Dumas, Impact of the gut microbiota on inflammation, obesity, and metabolic disease, *Genome Med.* **8** (2016) #42.

[12] J. H. Bramble, B. E. Hubbard, On a finite difference analogue of an elliptic problem which is neither diagonally dominant nor of nonnegative type, *J. Math. Phys.* **43** (1964) 117–132.

[13] R. Cambini, G. Gallo, M. G. Scutellá, Flows on hypergraphs, *Math. Prog.* **78** (1997) 195–217.

[14] N. S. Chandel, Glycolysis, *Cold Spring Harbor Persp. Biol.* **13** (2021) #a040535.

[15] T. Charoenrat, M. Ketudat-Cairns, H. Stendahl-Andersen, M. Jahic, S. O. Enfors, Oxygen-limited fed-batch process: An alternative control for Pichia pastoris recombinant protein processes, *Bioprocess Biosys. Engin.* **27** (2005) 399–406.

[16] Y. E. Cheah, J. D. Young, Isotopically nonstationary metabolic flux analysis (INST-MFA): Putting theory into practice, *Curr. Opinion Biotech.* **54** (2018) 80–87.

[17] S. Chen, S. An, R. Babazade, Y. Jung, Precise atom-to-atom mapping for organic reactions via human-in-the-loop machine learning, *Nat. Commun.* **15** (2024) #2250.

[18] T. Conway, The Entner–Doudoroff pathway: history, physiology and molecular biology, *FEMS Microbiol. Rev.* **9** (1992) 1–27.

[19] S. B. Crown, M. R. Antoniewicz, Parallel labeling experiments and metabolic flux analysis: Past, present and future methodologies, *Metab. Engin.* **16** (2013) 21–32.

[20] B. De Falco, F. Giannino, F. Carteni, S. Mazzoleni, D. H. Kim, Metabolic flux analysis: A comprehensive review on sample preparation, analytical techniques, data analysis, computational modelling, and main application areas, *RSC Adv.* **12** (2022) 25528–25548.

[21] H. C. Ehrlich, M. Rarey, Maximum common subgraph isomorphism algorithms and their applications in molecular science: a review, *WIREs Comp. Mol. Sci.* **1** (2011) 68–79.

[22] J. Eiler, J. Cesar, L. Chimiak, B. Dallas, K. Grice, J. Griep-Raming, D. Juchelka, N. Kitchen, M. Lloyd, A. Makarov, R. Robins, J. Schwieters, Analysis of molecular isotopic structures at high precision and accuracy by Orbitrap mass spectrometry, *Int. J. Mass Spectrometry* **422** (2017) 126–142.

[23] M. Feinberg, Complex balancing in general kinetic systems, *Arch. Rat. Mech. Anal.* **49** (1972) 187–194.

[24] C. Flamm, D. Merkle, P. F. Stadler, U. Thorsen, *Automatic Inference of graph transformation rules using the cyclic nature of chemical reactions*, 2016, comment: ICGT 2016 : 9th International Conference on Graph Transformation, extended technical report.

[25] T. Gallai, Maximum-Minimum Sätze über Graphen, *Acta Math. Acad. Sci. Hung.* **9** (1958) 395–434.

[26] C. Gauchotte-Lindsay, S. M. Turnbull, On-line high-precision carbon position-specific stable isotope analysis: A review, *Trends Anal. Chem.* **76** (2016) 115–125.

[27] M. E. González Laffitte, K. Weinbauer, T.-L. Phan, N. Beier, N. Domschke, C. Flamm, T. Gatter, D. Merkle, P. F. Stadler, Partial Imaginary transition state (ITS) graphs: A formal framework for research and analysis of atom-to-atom maps of unbalanced chemical reactions and their completions, *Symmetry* **16** (2024) #1217.

[28] T. Hasunuma, K. Harada, S.-I. Miyazawa, A. Kondo, E. Fukusaki, C. Miyake, Metabolic turnover analysis by a combination of in vivo 13C-labelling from 13CO2 and metabolic profiling with CE-MS/MS reveals rate-limiting steps of the C3 photosynthetic pathway in Nicotiana tabacum leaves, *J. Exp. Botany* **61** (2010) 1041–1051.

[29] M. K. Hellerstein, R. A. Neese, Mass isotopomer distribution analysis: a technique for measuring biosynthesis and turnover of polymers, *Am. J. Physiol.* **263** (1992) E988–E1001.

[30] D. Hershkowitz, Recent directions in matrix stability, *Lin. Algebra Appl.* **171** (1992) 161–186.

[31] D. W. Hoffman, C. Rasmussen, Position-specific carbon stable isotope ratios by proton NMR spectroscopy, *Anal. Chem.* **91** (2019) 15661–15669.

[32] F. Horn, R. Jackson, General mass action kinetics, *Arch. Rat. Mech. Anal.* **47** (1972) 81–116.

[33] M. E. Gonzalez Laffitte, N. Beier, N. Domschke, P. F. Stadler, Comparison of atom maps, *MATCH Commun. Math. Comput. Chem.* **90** (2023) 75–102.

[34] C. Jang, L. Chen, J. D. Rabinowitz, Metabolomics and Isotope tracing, *Cell* **173** (2018) 822–837.

[35] C. Jang, L. C. Chen, J. D. Rabinowitz, Metabolomics and isotope tracing, *Cell* **173** (2018) 822–837.

[36] C. Jochum, J. Gasteiger, I. Ugi, The principle of minimum chemical distance (PMCD), *Angew. Chem.* **19** (1980) 495–505.

[37] C. R. Johnson, Sufficient conditions for d-stability, *J. Econ. Th.* **9** (1974) 53–62.

[38] A. Lin, N. Dyubankova, T. I. Madzhidov, R. I. Nugmanov, J. Verhoeven, T. R. Gimadiev, V. A. Afonina, Z. Ibragimova, A. Rakhimbekova, P. Sidorov, A. Gedich, R. Suleymanov, R. Mukhametgaleev, J. Wegner, H. Ceulemans, A. Varnek, Atom-to-atom mapping: a benchmarking study of popular mapping algorithms and consensus strategies, *Mol. Inf.* **41** (2022) #2100138.

[39] F. Llaneras, J. Picó, Stoichiometric modelling of cell metabolism, *J. Biosci. Bioengin.* **105** (2008) 1–11.

[40] B. D. McKay, A. Piperno, Practical graph isomorphism, II, *J. Symb. Comput.* **60** (2014) 94–112.

[41] S. Müller, C. Flamm, P. F. Stadler, What makes a reaction network "chemical"? *J. Cheminf.* **14** (2022) #63.

[42] J. Nielsen, It is all about metabolic fluxes, *J. Bacteriol.* **185** (2003) 7031–7035.

[43] S. Noack, K. Nöh, M. Moch, M. Oldiges, W. Wiechert, Stationary versus non-stationary 13C-MFA: A comparison using a consistent dataset, *J. Biotech.* **154** (2011) 179–190.

[44] K. Nöh, W. Wiechert, The benefits of being transient: Isotope-based metabolic flux analysis at the short time scale, *Appl. Microbiol. Biotech.* **91** (2011) 1247–1265.

[45] R. Nugmanov, N. Dyubankova, A. Gedich, J. K. Wegner, Bidirectional graphormer for reactivity understanding: neural network trained to reaction atom-to-atom mapping task, *J. Chem. Inf. Mod.* **62** (2022) 3307–3315.

[46] J. D. Orth, I. Thiele, B. Ø. Palsson, What is flux balance analysis? *Nature Biotech.* **28** (2010) 245–248.

[47] N. Peekhaus, T. Conway, What's for dinner?: Entner-Doudoroff metabolism in *escherichia coli*, *J Bacteriol.* **180** (1998) 3495–502.

[48] T.-L. Phan, K. Weinbauer, M. E. G. Laffitte, Y. Pan, D. Merkle, J. L. Andersen, R. Fagerberg, C. Flamm, P. F. Stadler, `SynTemp`: Efficient extraction of graph-based reaction rules from large-scale reaction databases, *J. Chem. Inf. Model.* **65** (2025) 2882–2896.

[49] S. Qiu, A. Yang, H. Zeng, Flux balance analysis-based metabolic modeling of microbial secondary metabolism: Current status and outlook, *PLOS Comp. Biol.* **19** (2023) #e1011391.

[50] A. Rantanen, H. Maaheimo, E. Pitkänen, J. Rousu, E. Ukkonen, Equivalence of metabolite fragments and flow analysis of isotopomer distributions for flux estimation, in: C. Priami, G. Plotkin (Eds.), *Transactions on Computational Systems Biology VI*, Springer, Berlin, 2006, pp. 198–220.

[51] U. Roessner-Tunali, J. Liu, A. Leisse, I. Balbo, A. Perez-Melis, L. Willmitzer, A. R. Fernie, Kinetics of labelling of organic and amino acids in potato tubers by gas chromatography-mass spectrometry following incubation in (13)C labelled isotopes, *Plant J. Cell Mol. Biol.* **39** (2004) 668–679.

[52] A. H. Romano, T. Conway, Evolution of carbohydrate metabolic pathways, *Res. Microbiol.* **147** (1996) 448–455.

[53] S. Schuster, D. A. Fell, T. Dandekar, A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks, *Nature Biotech.* **18** (2000) 326–332.

[54] P. Schwaller, B. Hoover, J. L. Reymond, H. Strobelt, T. Laino, Extraction of organic chemistry grammar from unsupervised learning of chemical reactions, *Sci. Adv.* **7** (2021) #eabe4166.

[55] A. A. Shastri, J. A. Morgan, A transient isotopic labeling methodology for 13C metabolic flux analysis of photoautotrophic microorganisms, *Phytochem.* **68** (2007) 2302–2312.

[56] P. N. Shivakumar, K. H. Chew, A sufficient condition for nonvanishing of determinants, *Proc. Am. Math. Soc.* **43** (1974) 63–66.

[57] A. Stincone, A. Prigione, T. Cramer, M. M. C. Wamelink, K. Campbell, E. Cheung, V. Olin-Sandoval, N. M. Grüning, A. Krüger, M. Tauqeer Alam, M. A. Keller, M. Breitenbach, K. M. Brindle, J. D. Rabinowitz, M. Ralser, The return of metabolism: Biochemistry and physiology of the pentose phosphate pathway, *Biol. Rev. Cambridge Phil. Soc.* **90** (2015) 927–963.

[58] D. Talwar, C. G. Miller, J. Grossmann, L. Szyrwiel, T. Schwecke, V. Demichev, A.-M. Mikecin Drazic, A. Mayakonda, P. Lutsik, C. Veith, M. D. Milsom, K. Müller-Decker, M. Mülleder, M. Ralser, T. P. Dick, The GAPDH redox switch safeguards reductive capacity and enables survival of stressed tumour cells, *Nature Metabolism* **5** (2023) 660–676.

[59] O. Taussky, A recurring theorem on determinants, *Am. Math. Monthly* **56** (1949) 672–676.

[60] N. Tepper, E. Noor, D. Amador-Noguez, H. H. S, R. Milo, J. Rabinowitz, W. Liebermeister, T. Shlomi, Steady-state metabolite concentrations reflect a balance between maximizing enzyme efficiency and minimizing total metabolite load, *PLoS One* **8** (2013) #e75370.

[61] T. TeSlaa, M. Ralser, J. Fan, J. D. Rabinowitz, The pentose phosphate pathway in health and disease, *Nat. Metab.* **5** (2023) 1275–1289.

[62] L. Verde-Star, On linear matrix differential equations, *Adv. Appl. Math.* **39** (2007) 329–344.

[63] E. Vranová, D. Coman, W. Gruissem, Network analysis of the MVA and MEP pathways for isoprenoid synthesis, *Ann. Rev. Plant Biol.* **64** (2013) 665–700.

[64] W. Wiechert, K. Nöh, Isotopically non-stationary metabolic flux analysis: Complex yet highly informative, *Curr. Opinion Biotech.* **24** (2013) 979–986.

[65] W. Wiechert, C. Siefke, A. A. De Graaf, A. Marx, Bidirectional reaction steps in metabolic networks: II. Flux estimation and statistical analysis, *Biotech. Bioengin.* **55** (1997) 118–135.

[66] B. M. Woolston, S. Edgar, G. Stephanopoulos, Metabolic Engineering: Past and Future, *Ann. Rev. Chem. Biomol. Engin.* **4** (2013) 259–288.

[67] C. Wu, M. Guarnieri, W. Xiong, FreeFlux: A Python Package for Time-Efficient Isotopically Nonstationary Metabolic Flux Analysis, *ACS Synth. Biol.* **12** (2023) 2707–2714.

[68] J. D. Young, INCA: A computational platform for isotopically nonstationary metabolic flux analysis, *Bioinformatics* **30** (2014) 1333–1335.

[69] J. D. Young, J. L. Walther, M. R. Antoniewicz, H. Yoo, G. Stephanopoulos, An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis, *Biotech. Bioengin.* **99** (2008) 686–699.

[70] J. Zanghellini, D. E. Ruckerbauer, M. Hanscho, C. Jungreuthmayer, Elementary flux modes in a nutshell: Properties, calculation and applications, *Biotech. J.* **8** (2013) 1009–1016.

[71] Z. Zhao, A. Ten Pierick, L. de Jonge, J. J. Heijnen, S. A. Wahl, Substrate cycles in Penicillium chrysogenum quantified by isotopic non-stationary flux analysis, *Microb. Cell Fact.* **11** (2012) #140.

Appendix A Alternative Derivation of Simple ATGs

A simplified version for the raw atom transition graph of a reaction $r = (Q \to Q')$ can be derived by considering symmetries in the products only. To this end we start from $\zeta_v^1(u)$, defined as the set of atoms x' in the equivalence class $\zeta^{-1}(u)$ with an image in $\mathrm{orb}_{Q'}(y), y \in \zeta^{-1}(v)$ via the AAM φ and $\eta(v,u) = \sum_{\bar{u} \in \mathrm{orb}_{c_{\bar{u}}}} |\zeta_v^{-1}(u)|$ in Equ.(21):

Definition 13 (simplified raw atom transition graph)**.** Let $r = (Q \to Q')$ be a reaction with *AAM* $\varphi : V(Q) \to V(Q')$ and rATG $\tilde{T}_{QQ'}$. The simplified raw ATG is the graph $\tilde{S}_{QQ'}$ with vertex set $V(\tilde{S}_{QQ'}) \coloneqq V(\tilde{T}_{QQ'})$ and edge set $E(\tilde{S}_{QQ'}) \coloneqq \bigcup_{x \in V(Q)} E_{out}(x)$ with $E_{out}(x) \coloneqq \{(x,y) | y \in \rho'(\varphi(x))\}$, $\rho' \in \mathrm{Aut}(Q')$.

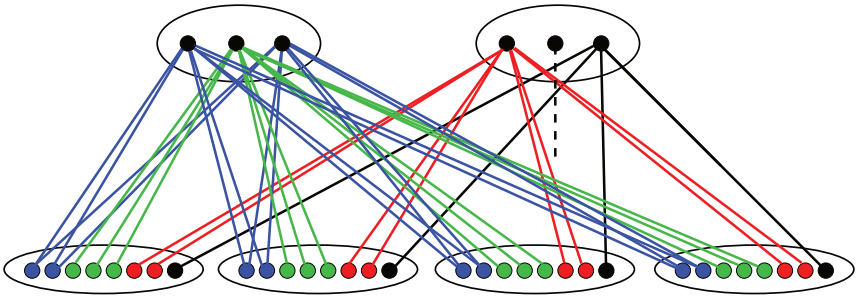

Figure 8. Simplified raw atom transition graph for the example AAM depicted in figure 3.

Figure 8 shows the simplified raw ATG for the example depicted in figure 3. Note that $\tilde{S}_{QQ'}$ is a subgraph of $\tilde{T}_{QQ'}$ since $\mathrm{Id} \in \mathrm{Aut}(Q)$. In addition and as in the raw ATG, the number of incoming edges is constant on $\mathrm{orb}_{Q'}(y)$, i. e. $E_{in}(y') = E_{in}(y'')$ for all $y', y'' \in \mathrm{orb}_{Q'}(y)$.

Defining a suitable edge-weight allows us to conserve valuations, such as the number of transferred atoms from $\mathrm{orb}_Q(x)$ to $\mathrm{orb}_{Q'}(y)$ for all $x \in V(Q), y \in V(Q')$. For a (simplified) raw ATG $\tilde{S}_{QQ'}$ or $\tilde{T}_{QQ'}$ we therefore

define

$$E_{xy}^{\tilde{S}/\tilde{T}} := \{(x', y') \in E(\tilde{S}/\tilde{T}_{QQ'}) | x' \in \mathrm{orb}_Q(x), y' \in \mathrm{orb}_{Q'}(y)\} \tag{34}$$

as the set of edges between $\mathrm{orb}_Q(x)$ and $\mathrm{orb}_{Q'}(y)$.

Lemma 14. *Let $r = (Q \to Q')$ be a reaction with simplified raw ATG $\tilde{S}_{QQ'}$ and $h_{\tilde{S}} : E(\tilde{S}_{QQ'}) \to \mathbb{R}, h_{\tilde{S}}(y, x) = |\mathrm{orb}_{Q'}(y)|^{-1}$. Then:*

$$\eta(y, x) = \sum_{(x',y') \in E_{xy}^{\tilde{S}}} h_{\tilde{S}}(y', x') \tag{35}$$

Proof. Since the weights of edges between $\mathrm{orb}_Q(x)$ and $\mathrm{orb}_{Q'}(y)$ are identical we only have to show that $|E_{xy}^{\tilde{S}}| = \eta(y, x) \cdot |\mathrm{orb}_{Q'}(y)|$. Let, therefore, $u := \zeta(x)$. For $x' \in \mathrm{orb}_Q(x)$, $y' \in \mathrm{orb}_{Q'}(y)$ we have $(x', y') \in E(\tilde{S}_{QQ'}) \Leftrightarrow x' \in \zeta_y^{-1}(u')$ for $u' \in \mathrm{orb}_{c_u}(u)$ as defined in Equ. 18. Then

$$\begin{aligned} \left|E_{xy}^{\tilde{S}}\right| &= \sum_{y' \in \mathrm{orb}_{Q'}(y)} \sum_{u' \in \mathrm{orb}_{c_u}(u)} \sum_{x' \in \zeta_y^{-1}(u')} 1 \\ &= \sum_{y' \in \mathrm{orb}_{Q'}(y)} \sum_{u' \in \mathrm{orb}_{c_u}(u)} |\zeta_y^{-1}(u')| \\ &\overset{Equ.(22),\ *}{=} |\mathrm{orb}_{Q'}(y)| \cdot \eta(y, x) \end{aligned} \tag{36}$$

■

Lemma 14 guarantees that the number of atoms transferred from $\mathrm{orb}_Q(x)$ to $\mathrm{orb}_{Q'}(y)$ in reaction r is conserved also in the simplified version of the raw atom transition graph. In addition, exactly $\eta(y, x)$ vertices in $\tilde{S}_{QQ'}$ map from $\mathrm{orb}_Q(x)$ to one particular $y' \in \mathrm{orb}_{Q'}(y)$.

We proceed by computing valuations on the simplified raw ATG and show that they coincide with valuations derived from the construction of raw ATGs.

Lemma 15. *Let $r = (Q \to Q')$ be a reaction with simplified raw ATG $\tilde{S}_{QQ'}$, $\lambda : V(Q) \to \mathbb{R}$ a valuation on Q that is constant on the orbits of $\mathrm{Aut}(Q)$, and $h_{\tilde{S}} : E(\tilde{S}_{QQ'}) \to \mathbb{R}, h_{\tilde{S}}((x, y)) = |\mathrm{orb}_{Q'}(y)|^{-1}$. Then:*

$$\lambda'(y) = \sum_{(x,y) \in E_{in}^{\tilde{S}}(y)} h_{\tilde{S}}(y, x) \cdot \lambda(x) \tag{37}$$

$$\sum_{(x,y) \in E_{in}^{\tilde{S}}(y)} h_{\tilde{S}}(y, x) \cdot \lambda(x) = \sum_{x \in V(Q)} \lambda(x) \cdot \frac{\eta(x, y)}{|\mathrm{orb}_Q(x)| \cdot |\mathrm{orb}_{Q'}(y)|} \tag{38}$$

Proof. We start with the first statement. Recall by Equ. (16) we have

$$\begin{aligned} \lambda'(Y) &= \sum_{A\in\Theta(Q)} \frac{1}{|Y|} \sum_{y'\in Y} \sum_{x\in A} \lambda(x)\cdot \mathbb{I}[y'=\varphi(x)] \\ \Rightarrow \lambda'(y) &= \sum_{A\in\Theta(Q)} \sum_{x\in A} \sum_{y'\in \mathrm{orb}_{Q'}(y)} \lambda(x)\cdot \frac{1}{|\,\mathrm{orb}_{Q'}(y)|}\mathbb{I}[y'=\varphi(x)] \end{aligned} \tag{39}$$

By definition of $\tilde{S}_{QQ'}$ we have $(x,y)\in E(\tilde{S}_{QQ'})$ if and only if there is an $y'\in \mathrm{orb}_{Q'}(y) = Y : y' = \varphi(x)$. Hence, we recover all edges pointing into $\mathrm{orb}_{Q'}(y')$ because $E_{in}(y) = \{(x,y') \mid x\in A\subseteq\Theta(Q), y'\in \mathrm{orb}_{Q'}(y), y' = \varphi(x)\}$. This yields

$$\begin{aligned} \lambda'(y) &= \sum_{A\in\Theta(Q)} \sum_{x\in A} \sum_{y'\in \mathrm{orb}_{Q'}(y)} \lambda(x)\cdot \frac{1}{|\,\mathrm{orb}_{Q'}(y)|}\mathbb{I}[y'=\varphi(x)] \\ &= \sum_{(x,y)\in E^{\tilde{S}}_{in}(y)} h_{\tilde{S}}(y,x)\cdot\lambda(x) \end{aligned} \tag{40}$$

The second part follows from the computation above and Lemma 9. ■

Let us now turn to the construction of simplified ATG from the simplified raw ATG. To this end, we restrict our construction for simplified raw ATGs to the non-isomorphic connected components of Q and Q'. As a result, we obtain a sparse version of the ATG.

Definition 14. Let $r = (Q\to Q')$ be a reaction with AAM $\varphi : V(Q)\to V(Q')$ and simplified raw ATG $\tilde{S}_{QQ'}$. The *alternative simplified atom transition graph* $\mathfrak{S}_{QQ'}$ is the quotient graph of $\tilde{S}_{QQ'}$ w.r.t. the equivalence relation ζ^{-1} defined by the isomorphic connected components of Q and Q', respectively.

In the following we aim to show the equivalence of the upper construction with the one of simplified ATGs of definition 10. In order to accomplish this goal, we need to show to that both graphs are identical, in terms of vertices and edges, and that valuations on $V(Q'_\circ)$ are preserved.

Lemma 16. *Let $r = (Q\to Q')$ be a reaction with raw simplified, simplified and alternative simplified ATGs $\tilde{S}_{QQ'}, S_{QQ'}$ and $\mathfrak{S}_{QQ'}$, respectively. Then $S_{QQ'} = \mathfrak{S}_{QQ'}$.*

Proof. By construction $V(S_{QQ'}) = V(\mathfrak{S}_{QQ'})$. We denote by $E^{\tilde{S}} \coloneqq E(\tilde{S}_{QQ'}), E^S \coloneqq E(S_{QQ'})$, and $E(\mathfrak{S}) \coloneqq E(\mathfrak{S}_{QQ'})$. We have $(u,v)\in E^S \Leftrightarrow \zeta_v^{-1}(u)\neq\emptyset \Leftrightarrow \exists x'\in\zeta^{-1}(u) \mid \exists\bar{v}\in \mathrm{orb}_{c'_v}(v) : \varphi(x')\in\zeta^{-1}(\bar{v})$. Hence, we

have $(x', y') \in E^{\tilde{S}}$ for all $x' \in \zeta_u^{-1}(v)$ and $y' \in \zeta^{-1}(\bar{v}) : \bar{v} \in \mathrm{orb}_{c'_v}(v)$. This yields for the quotient graph that $(u, v) \in E^{\mathfrak{S}}$. We conclude $E^S \subseteq E^{\mathfrak{S}}$. On the other hand, $(u, v) \in E^{\mathfrak{S}}$ implies that there exists $x' \in \zeta^{-1}(u), y' \in \zeta^{-1}(v) : (x', y') \in E^{\tilde{S}}$ and in particular that $\zeta_u^{-1}(v) \neq \emptyset$, which directly yields $(u, v) \in E^S$. ∎

By lemma 10 we directly obtain that also $\mathfrak{S}_{QQ'}$ is a subgraph of $T_{QQ'}$. It remains to show that the valuations are conserved on $\mathfrak{S}_{QQ'}$. For this purpose we need to assign appropriate edge weights, which we will derive subsequently. In the following we will continue to use $\mathfrak{S}_{QQ'}$ instead of $S_{QQ'}$ to have a consistent notation for the edge weights. The main idea will be to maintain the number of atoms transfered from $\mathrm{orb}_Q(x)$ and $\mathrm{orb}_{Q'}(y)$ for each $v \in V(Q), y \in V(Q')$. As before, let $x \in V(Q), y \in V(Q')$ be fixed and $u := \zeta(x), v := \zeta(y)$. We observe that in $\mathfrak{S}_{QQ'}$, vertices $y' \in \mathrm{orb}_{Q'}(y)$ from $\tilde{S}_{QQ'}$ are merged to equivalent classes. Therefore, the number of edges received from $\mathrm{orb}_Q(x)$ by $\mathrm{orb}_{c'}(\zeta(y))$ in $\mathfrak{S}_{QQ'}$ is given by:

$$\frac{1}{s^+_{c'r}} \cdot \left| E^{\tilde{S}}_{xy} \right| \overset{Lem.14}{=} \frac{|\,\mathrm{orb}_{Q'}(y)|}{s^+_{c'r}} \cdot \eta(y, x) = |\,\mathrm{orb}_{c'_{\zeta(y)}}(\zeta(y))| \cdot \eta(y, x) \quad (41)$$

Thus, an $x' \in \mathrm{orb}_Q(x)$ adds at most $|\,\mathrm{orb}_{c'_{\zeta(y)}}(\zeta(y))|$ edges. However, in $\mathfrak{S}_{QQ'}$ only those equivalence classes with $\zeta_v^{-1}(\zeta(x')) \neq \emptyset$ contribute to the pool of transition edges between $\mathrm{orb}_{c_{\zeta(x)}}(\zeta(x))$ and $\mathrm{orb}_{c'_{\zeta(y)}}(\zeta(y))$. To this end we define

$$Z_v := \{\zeta^{-1}(u) | \zeta_v^{-1}(u) \neq \emptyset\}. \quad (42)$$

Note that, since φ is a bijection, we have $|Z_v| > 0$ for all $v \in V(Q'_\circ)$. Thus, with $\zeta(x) = u, \zeta(y) = v$, and $\eta(y, x) = \eta(v, u)$ the number of edges mapping from $\mathrm{orb}_{c_u}(u)$ to $\mathrm{orb}_{c'_v}(v)$ in $\mathfrak{S}_{QQ'}$ is given by

$$\eta(v, u) \cdot \mathrm{orb}_{c'_v}(v) \cdot \frac{|Z_v|}{\eta(v, u)} = \mathrm{orb}_{c'_v}(v) \cdot |Z_v| \quad (43)$$

Lemma 17. *Let $r = (Q \to Q')$ be a reaction with simplified atom transition graph $\mathfrak{S}_{QQ'}$ and set $E^{\mathfrak{S}}_{uv} = \{(u', v) | (u' \in \mathrm{orb}_{c_u}(u))\}$ for $u \in V(Q_\circ)$, and $Z_v := \{\zeta^{-1}(u) | \zeta_v^{-1}(u) \neq \emptyset\}$. Then*

$$\sum_{(u', v) \in E^{\mathfrak{S}}_{uv}} \frac{1}{|Z_v|} = 1 \quad (44)$$

Proof. By construction of $\mathfrak{S}_{QQ'}$, each $v' \in \mathrm{orb}_{c'_v}(v)$ receives the same

number of edges form $\mathrm{orb}_{c_u}(u)$. Hence

$$\left|E^{\mathfrak{S}}_{uv}\right| = |\,\mathrm{orb}_{c'_v}(v)| \cdot \frac{|Z_v|}{|\,\mathrm{orb}_{c'_v}(v)|} = |Z_v| \tag{45}$$

From the discussion above we obtain that the number of edges between $\mathrm{orb}_{c_u}(u)$ and $\mathrm{orb}_{c'_v}(v)$ is $|\,\mathrm{orb}_{c'_v}(v)|\cdot|Z_v|$. Combining these two observations yields Equ. (44). ■

Defining the weight of edges from $\mathrm{orb}_{c_u}(u)$ to $\mathrm{orb}_{c'_v}(v)$ in $\mathfrak{S}_{QQ'}$ by means of the map $h_{\mathfrak{S}} : E(\mathfrak{S}_{QQ'}) \to \mathbb{R}$ with

$$h_{\mathfrak{S}}(v,u) \coloneqq \frac{\eta(v,u)}{|Z_v| \cdot |\,\mathrm{orb}_{c'_v}(v)|} \tag{46}$$

therefore maintains the number of transferred atoms. These weights represent an analogous definition to the weights assigned to transition edges in $T_{QQ'}$ as we only had to exchange the size of $\mathrm{orb}_{c_u}(u)$ by $|Z_v|$ which is reflected by the different sizes of starting vertices contributing to the transition edge pool. We will now see that engaging h_S as edge weights, also valuations are conserved in $S_{QQ'}$ as they are conserved in $T_{QQ'}$.

Lemma 18. *Let $r = (Q \to Q')$ be a reaction with atom transition graph $T_{QQ'}$ and simplified ATG $\mathfrak{S}_{QQ'}$ with edge sets $E^{\mathfrak{S}} := E(\mathfrak{S}_{QQ'})$ and $E^T := E(T_{QQ'})$ Let $\lambda : V(Q) \to \mathbb{R}$ be a valuation that is constant on the orbits of $\mathrm{Aut}(Q)$ and $\lambda'(v) := \sum_{(u,v)\in E^T_{in}(v)} s^+_{c'_v r} \cdot h(v,u) \cdot \lambda(u)$ with $h(v,u)$ as defined in Equ. (12). Then:*

$$\lambda'(v) = \sum_{(u,v)\in E^{\mathfrak{S}}_{in}(v)} h_{\mathfrak{S}}(v,u) \cdot \lambda(u) \tag{47}$$

for all $v \in V(Q')$.

Proof. We aim to show that

$$\sum_{(u,v)\in E^T_{in}(v)} s^+_{c_v} r \cdot h(v,u) \cdot \lambda(u) = \sum_{(u,v)\in E^{\mathfrak{S}}_{in}(v)} h_{\mathfrak{S}}(v,u) \cdot \lambda(u)\,.$$

Recalling that the orbits of $\mathrm{Aut}(Q)$ form a partition of $V(Q)$ into subsets $A \in \Theta(Q)$ and substituting the definitions of $h(v,u)$ and $h_{\mathfrak{S}}(v,u)$ yields

$$\sum_{A\in\Theta(Q)}\sum_{x\in A}\sum_{(\zeta(x),v)\in E_{in}^{T}(v)}\frac{\eta(v,\zeta(x))}{|\operatorname{orb}_{c_{\zeta(x)}}(\zeta(x))|}\lambda(\zeta(x))$$
$$=\sum_{A\in\Theta(Q)}\sum_{x\in A}\sum_{(\zeta(x),v)\in E_{in}^{\mathfrak{S}}(v)}\frac{\eta(v,\zeta(x))}{|Z_v|}\lambda(\zeta(x))\,.$$

To see that this is indeed an identity we observe that Equ. (44) can be rewritten in the form

$$\sum_{x\in A}\sum_{(\zeta(x),v)\in E_{in}^{T}(v)}\frac{1}{|\operatorname{orb}_{c_{\zeta(x)}}(\zeta(x))|}=\sum_{x\in A}\sum_{(\zeta(x),v)\in E_{in}^{\mathfrak{S}}(v)}\frac{1}{|Z_v|}$$

by observing that $\eta(v,\zeta(x))$ for fixed v and $\lambda(\zeta(x))$ are constant on the orbit $\operatorname{orb}_{c_{\zeta(x)}}(\zeta(x))$ and using $|E_{in}^{T}(v)| = |\operatorname{orb}_{c_{\zeta(x)}}(\zeta(x))|$ as well as $|E_{in}^{S}(v)| = |Z_v|$. ■

Finally, we show that valuations derived via h^* and $h_{\mathfrak{S}}$ only differ by their respective normalization.

Lemma 19. *Let $r = (Q \to Q')$ a reaction with simplified ATG $S_{QQ'}$. Then*

$$\sum_{u\in V(Q_\circ)} s^{+}_{c_v r}\cdot h^*(v,u)\cdot\lambda(u)=\sum_{(u,v)\in E_{in}^{\mathfrak{S}}(v)} h_{\mathfrak{S}}(v,u)\cdot\lambda(u) \tag{48}$$

for all $v \in V(Q_\circ)$.

Proof. The proof is a consequence of lemmas 9 and 18, and Equ. (17). ■