DIHYPERGRAPHS

Philipp-Jens Ostermeier, Jürgen Jost

Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

email: oster@mis.mpg.de, jjost@mis.mpg.de

MAX-PLANCK-GESELLSCHAFT

Abstract

The topology of metabolic networks [2] has often been studied using reaction-reaction or substrate digraphs [3],[1]. Spurious effects of such projections of the original dihypergraph have also been extensively commented upon. We study here how such projections can be improved at the expense of introducing weights on the arcs.

Definitions

- a *directed hypergraph* or *dihypergraph* H is a tuple $(V_H, \mathscr{E}_H)$ with
  - V_H is a set of vertices
  - each $e \in \mathscr{E}_H$ is an ordered tuple $(t(e), h(e))$ such that $t(e)$ is the tail and $h(e)$ the head of e, both are multisets on V_H
- $w_m(b)$ is the number of b being in the multiset m
- $|m|$ is the cardinality of a multiset m
- further we call a set with assigned weights to the elements a weighted set
- if m is a weighted set then m^s denotes the maximal underlying set such that we assign weight 1 to every element of m
- a *directed graph* or *digraph* D is a tuple (V_D, E_D) with
  - V_D is a set of vertices
  - the weighted arcs are of the form $e = ((a,b),(w_e^-(a), w_e^+(b)))$ with $w_e^-(a) \in \mathbf{Q}$ is the weight of $a \in V_D$ and $w_e^+(b) \in \mathbf{Q}$ is the weight of $b \in V_D$, we will simply write $(w_e^-(a)a, w_e^+(b)b)$
- if $e = (a,b)$ is an weighted (hyper-) arc then e^s denotes (a^s, b^s)

If all weights are 1 in a hyperdigraph we will not draw them in the pictures.

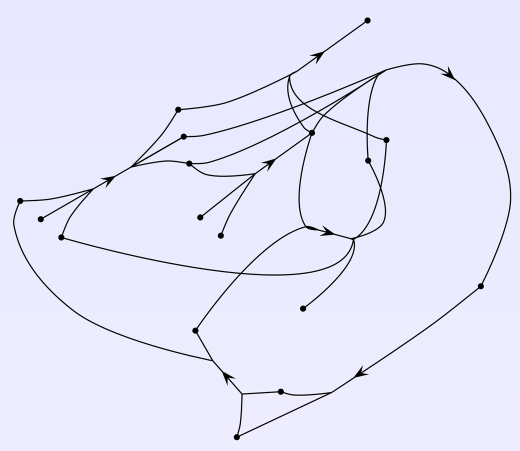

FIGURE 1: A dihypergraph on seventeen nodes and seven hyperarcs.

Substrate Digraph

Unweighted (with weight 1 on every vertex) substrate digraph:

The *substrate digraph* [4] $S(H)$ of the dihypergraph H is the digraph which considers V_H as vertices. There is an arc from $a \in V_H$ to $b \in V_H$ if there is an $e \in \mathscr{E}_H$ with $a \in t(e)$ and $b \in h(e)$.

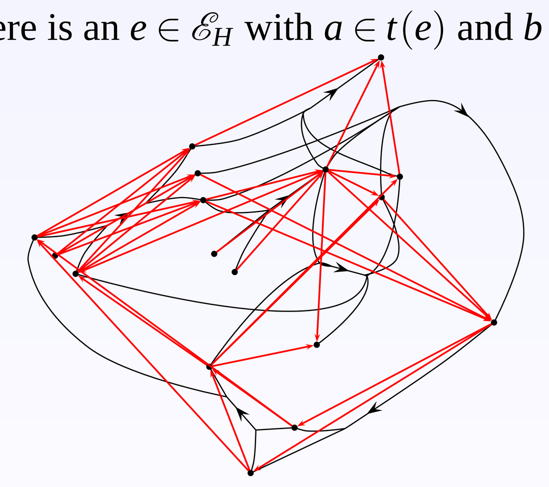

FIGURE 2: Substrate Digraph of the dihypergraph in FIGURE 1.

One advantage of the substrate digraph $S(H)$ over a bipartite representation $B(H)$ of a dihypergraph H is that it has much fewer dipaths and vertices. Moreover, $B(H)$ contains the hyperarcs of H as vertices.

The number of arcs of the substrate digraph of a hyperarc equals the number of 2-paths of the bipartite representation of this arc. An example is a hyperarc e with $t(e) \cap h(e) = \emptyset$, $S(e)$ has $|t(e^s)||h(e^s)|$ dipaths and $|t(e^s)| + |h(e^s)|$ vertices. $B(e)$ has $|t(e^s)||h(e^s)| + |t(e^s)| + |h(e^s)|$ dipaths and $|t(e^s)| + |h(e^s)| + 1$ vertices. It makes sense to use the substrate digraph for analyzing biological networks to find short paths from educts and products. The common construction of a substrate digraph loses most informations about allocating educts or products to reactions. We want to define a weighted model of the substrate digraph to keep a biochemically relevant part of those allocations. Unfortunately, different dihypergraphs may have the same substrate digraph $S(H)$ such that we cannot always find a unique mapping that sends $S(H)$ to H. In the following we give a weight construction on such substrate digraphs to keep weights better and to reduce the number of weighted dihypergraphs that determine such a digraph.

Weighted Substrate Digraph

First some basic definitions:

For $e \in \mathscr{E}_H, a \in t(e), b \in h(e)$:

- $(i_1(e), i_2(e)) = \begin{cases} (0,0) & \text{if} |t(e)| = |h(e)| \\ (0,1) & \text{if} |t(e)| > |h(e)| \\ (1,0) & \text{if} |t(e)| < |h(e)| \end{cases}$
- $c_e^-(a) := (2 \cdot |t(e)| \cdot |h(e^s)|)^{-1} \cdot w_{t(e)}(a)$

  $w_e^-(a) := (-1)^{i_1(e^s)} \cdot \left(c_e^-(a) + |t(e)|\right)$
- $c_e^+(a) := (2 \cdot |h(e)| \cdot |t(e^s)|)^{-1} \cdot w_{h(e)}(a)$

  $w_e^+(a) := (-1)^{i_2(e^s)} \cdot \left(c_e^+(a) + |h(e)|\right)$

$D : \mathscr{H} \to \mathscr{D}$ maps dihypergraphs to directed graphs:

$D(H)$

- $V_{D(H)} = V_H$
- $(f_1 \cdot a, f_2 \cdot b)$ is an arc in $E_{D(H)}$ if there is a hyperarc $e \in \mathscr{E}_H$ with $a \in t(e), b \in h(e)$ and $f_1 = w_e^-(a)$ and $f_2 = w_e^+(b)$

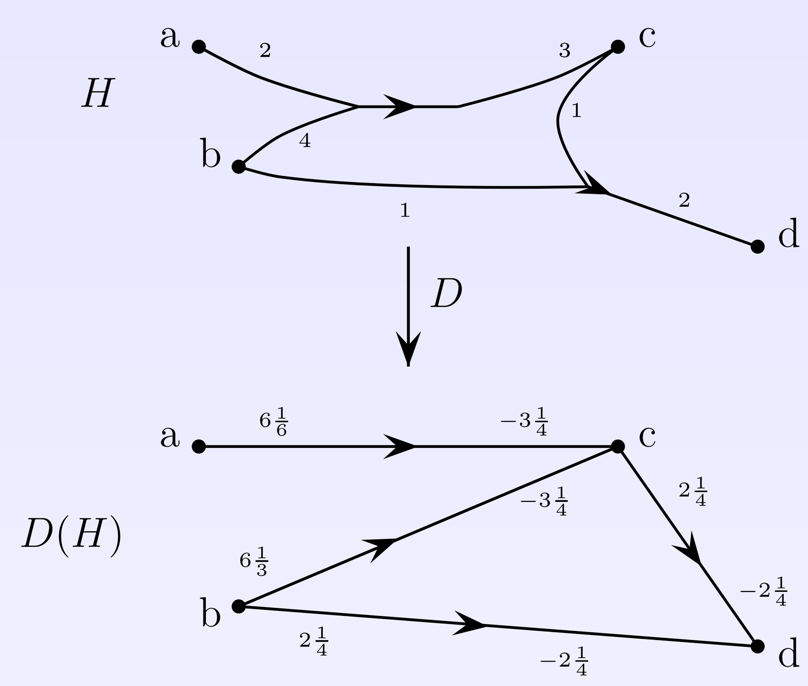

FIGURE 3: The weighted digraph $D(H)$ is uniquely determined by the weighted directed hypergraph H. The two weighted hyperarcs induce exactly 4 weighted arcs.

Some Properties:

For a subdihypergraph (V_e, e) of $H = (V_H, \mathscr{E}_H)$ induced by $e \in \mathscr{E}_H$ follows :

1. $\lfloor |w_e^-(a)| \rfloor = |t(e)|$ and $\lfloor |w_e^+(b)| \rfloor = |h(e)|$ for $a \in t(e), b \in h(e)$
2. $|\sum_{a \in t(e)} \left((|w_e^-(a)| - \lfloor |w_e^-(a)| \rfloor) \cdot |h(e^s)| \right)| = 0.5$
3. $|\sum_{a \in h(e)} \left(|(w_e^+(a)| - \lfloor |w_e^+(a)| \rfloor) \cdot |t(e^s)| \right)| = 0.5$
4. e^s induces a $\overrightarrow{K_{m,n}}$ in $D(H)$

In the following we restrict us to directed graphs in which every arc e is contained in an arc set $E' = \bigcup_{i=1,\dots,n} \bigcup_{j=1,\dots,m} \{e_{ij}\}$ with $e_{ij}^s = (a_i, b_j)$ for which the following conditions $1. - 9.$ are satisfied:

1. for fixed i: $w_{e_{ik}}^-(a_i) = w_{e_{il}}^-(a_i)$, $k, l \in \{1, \dots, m\}$
2. for fixed j: $w_{e_{kj}}^+(b_j) = w_{e_{lj}}^+(b_j)$, $k, l \in \{1, \dots, n\}$
3. $\lfloor |w_{e_{ij}}^-(a_i)| \rfloor = \lfloor |w_{e_{fg}}^-(a_f)| \rfloor \forall i, f \in \{1, \dots, n\} \forall j, g \in \{1, \dots, m\}$
4. $\lfloor |w_{e_{ij}}^+(b_j)| \rfloor = \lfloor |w_{e_{fg}}^+(b_g)| \rfloor \forall i, f \in \{1, \dots, n\} \forall j, g \in \{1, \dots, m\}$
5. $\sum_{i=1}^n \sum_{j=1}^m (|w_{e_{ij}}^-(a_i)| - \lfloor |w_{e_{ij}}^-(a_i)| \rfloor) = 0.5$
6. $\sum_{i=1}^n \sum_{j=1}^m (|w_{e_{ij}}^+(b_j)| - \lfloor |w_{e_{ij}}^+(b_j)| \rfloor) = 0.5$
7. if $m < n \Rightarrow w_{e_{ij}}^-(a_i) > 0, w_{e_{ij}}^+(b_j) < 0 \forall i, j$
8. if $m > n \Rightarrow w_{e_{ij}}^-(a_i) < 0, w_{e_{ij}}^+(b_j) > 0 \forall i, j$
9. if $m = n \Rightarrow w_{e_{ij}}^-(a_i) > 0, w_{e_{ij}}^+(b_j) > 0 \forall i, j$

The set of digraphs which satify these conditions will be denoted as $\mathscr{D}'$.
Define $D^* : \mathscr{D}' \to \mathscr{H}$ that maps weighted digraphs of $\mathscr{D}'$ to weighted dihypergraphs as follows:
D^* maps a set like E' to the hyperarc E with educts $a_1, a_2, \dots, a_n$ and weights $w_{t(e)}(a_i) = 2m(\lfloor |w_{e_{ij}}^-(a_i)| \rfloor - |w_{e_{ij}}^-(a_i)|) \lfloor |w_{e_{ij}}^-(a_i)| \rfloor$ for fixed j and products $b_1, \dots, b_m$ with weights $w_{h(e)}(b_j) = 2n(\lfloor |w_{e_{ij}}^+(b_j)| \rfloor - |w_{e_{ij}}^+(b_j)|) \lfloor |w_{e_{ij}}^+(b_j)| \rfloor$ with fixed i in the arc set $\mathscr{E}_H$. If the set $E' \subseteq E_D$ is not such a set it is mapped to the empty arc.

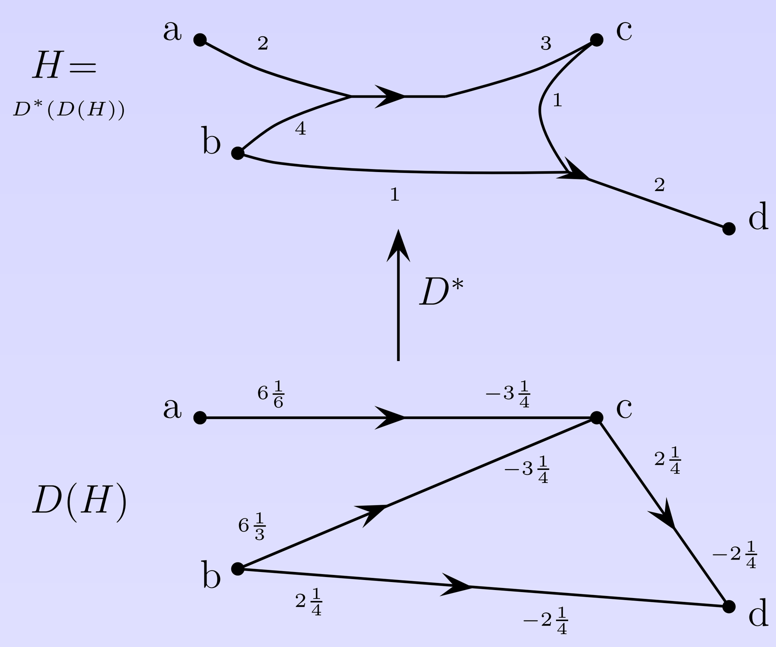

FIGURE 4: Weighted Substrate Digraph induces a unique dihypergraph such that H equals $D^*(D(H))$.

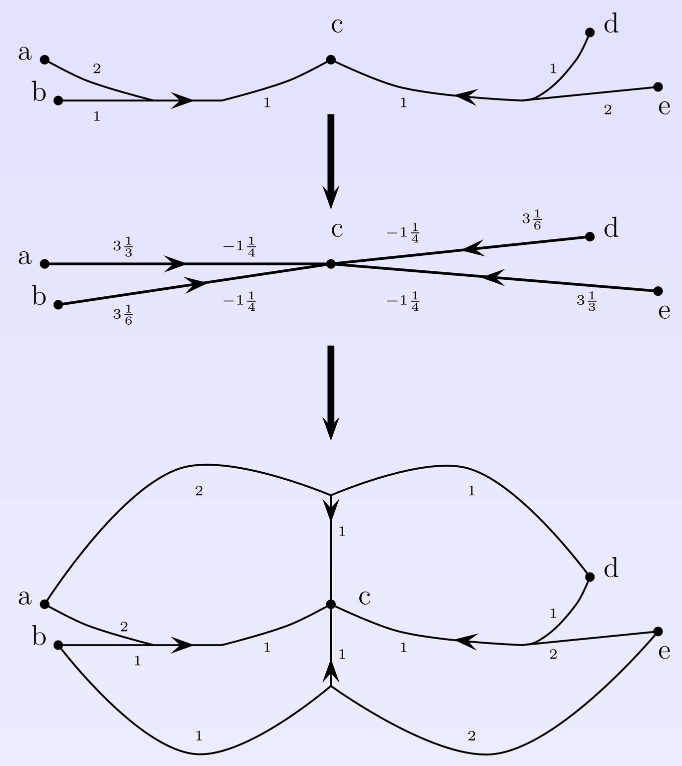

FIGURE 5: In this figure H has two dihyperarcs with the same head and the same weight distribution so is H a subgraph of $D^*(D(H))$.

Conclusions

- Given an ordered arc set which induces a weighted directed circuit $C \subseteq E_{D(H)}$. If C does not fulfill the following condition for every vertex

  The integer part of the absolute value of the in-going weight is greater or equal than the integer part of the absolute value of the out-going weight.

  then the subdihypergraph of H inducing E_C is not autocatalytic (every tail of an hyperarc is contained in another head of a graph).
- H is a directed subhypergraph of $D^*(D(H))$.
- Assume a dihypergraph H with weight 1 everywhere. For $S(H)$ exists a dihypergraph H' with $S(H) = S(H')$ such that for all complete bipartite subdigraphs exist a dihyperarc in H' with the same vertex set. If $S(H)$ contains a complete bipartite subdigraph with more than two vertices then there are three directed directed subhypergraphs of H' that also induce $S(H)$. For $D^*(D(H))$ is $D^*(D(H)) \subseteq H'$.

Acknowledgements:

I thank P.-Y. Bourguignon, Konstantin Klemm, Bruno Pace, Johannes Rauh, Peter Stadler and especially the VolkswagenStiftung.

References

[1] Ernesto Estrada and Juan A. Rodriguez-Velazquez. Complex networks as hypergraphs, 2005.

[2] H Jeong, B Tombor, R Albert, Z N Oltval, and A-L Barabasi. The large-scale organization of metabolic networks. 407:651–654, October 2000.

[3] Steffen Klamt, Utz-Uwe Haus, and Fabian Theis. Hypergraphs and cellular networks. *PLoS Comput Biol*, 5(5):e1000385, 05 2009.

[4] Andreas Wagner and David Fell. The small world inside large metabolic networks, 2000.