Ordered Colourings of Graphs

JAN VAN DEN HEUVEL

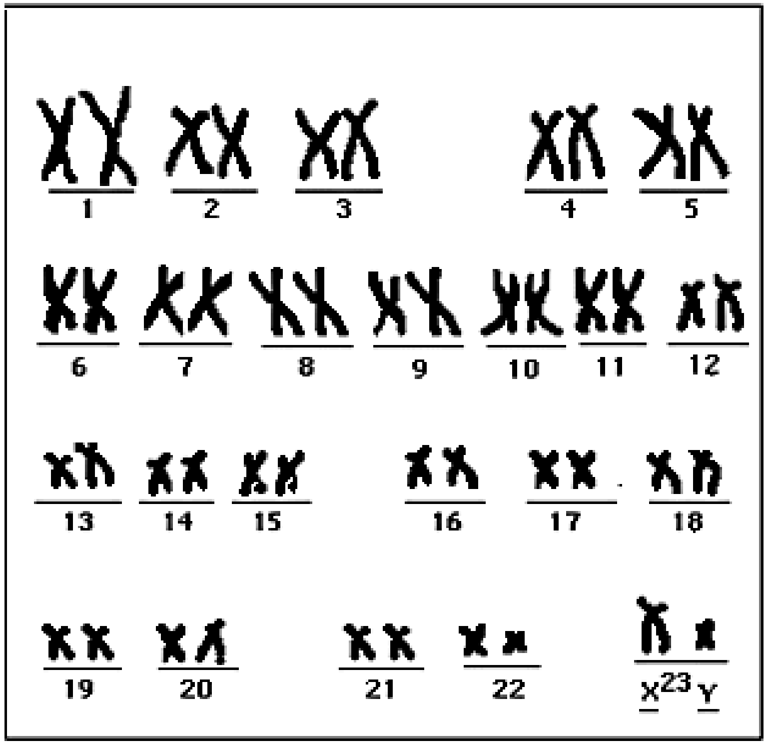
joint work with ARVIND GUPTA, JÁN MAŇUCH
LADISLAV STACHO & XIAOHONG ZHAO
(Simon Fraser University, Canada)

Centre for Discrete and Applicable Mathematics
Department of Mathematics

London School of Economics and Political Science

Inspired by a problem in mathematical biology

- **genotype** : the genetic information of one individual
  - stored on **chromosomes**
- most organisms have **two copies** of each chromosome



Inspired by a problem in mathematical biology

- **genotype** : the genetic information of one individual
  - stored on **chromosomes**
- most organisms have **two copies** of each chromosome
  - one from each parent
  - these two copies need **not be identical**
  - the genetic make-up of one chromosome is called a **haplotype**
- the genotype has the **combined information** of the two haplotypes in some kind of "**mixed format**"
- determining the haplotypes in a laboratory is **much harder** than determining the genotype

Genotype versus haplotypes

-
  - information in the genotype can be one of **0**, **1**, **2**
  - for the haplotype it can be one of **0**, **1**
- interpretation :
  - — **0** — → — **0** — / — **0** —
  - — **1** — → — **1** — / — **1** —
  - — **2** — → — **0** — / — **1** —

Main problem

- — **2** — **2** — → — **0** — **0** — / — **1** — **1** — **or** — **0** — **1** — / — **1** — **0** — **?**

Haplotype inferring

- **Input**: the different **variants of genotypes** appearing in some population

  Output: for each genotype: two corresponding **haplotypes**

- satisfying some **rules** (usually biologically inspired)

  - **e.g.**: the fewer variants of haplotypes required, the better

    $\{$ **02**, **20**, **22** $\}$ could be $\left\{ \begin{matrix} \mathbf{00} \\ \mathbf{01} \end{matrix}, \begin{matrix} \mathbf{00} \\ \mathbf{10} \end{matrix}, \begin{matrix} \mathbf{00} \\ \mathbf{11} \end{matrix} \right\}$

    or $\left\{ \begin{matrix} \mathbf{00} \\ \mathbf{01} \end{matrix}, \begin{matrix} \mathbf{00} \\ \mathbf{10} \end{matrix}, \begin{matrix} \mathbf{01} \\ \mathbf{10} \end{matrix} \right\}$

    2nd choice preferable (only three different haplotypes)

Rules from genetic assumptions

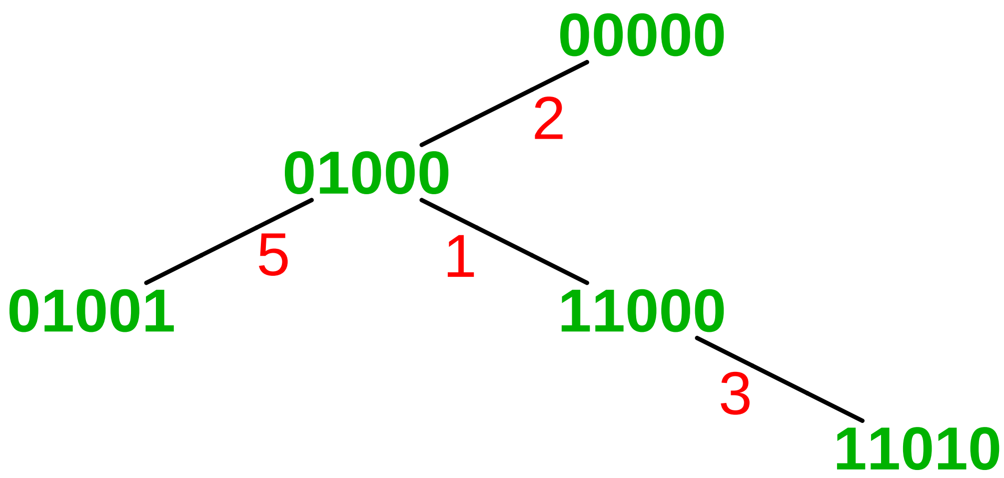
- initial assumptions :
  - **all** haplotypes originate from **one original** one
  - **variants** appeared because of **mutations**

    but only **small chance** that a mutation occurs at a particular place on a chromosome

- so try to obtain a collection of haplotypes requiring **minimum number of mutations**

An even stronger genetic requirement

- the proposed collection of haplotypes should should fit in a **Perfect Phylogenetic Tree** :
  - rooted tree
  - - each node labelled with a proposed haplotype
    - each haplotype appears as the label of one node
  - - each edge labelled by sites in which the ends differ
    - each site occurs at most once as an edge label



Perfect Phylogeny Haplotype Problem

- **Input** : the different **variants of genotypes** appearing in some population

  Question : is there a collection of haplotypes such that

  - each genotype has two corresponding haplotypes,
  - the haplotypes fit in a Perfect Phylogenetic Tree,
  - all haplotypes are used in some genotype ?

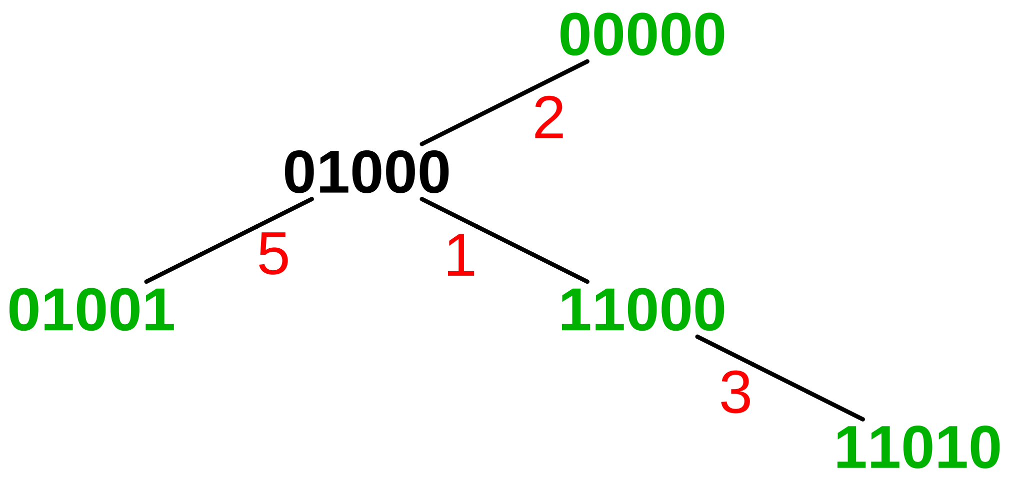
Theorem (Gusfield, 2002)

- *this can be solved in **polynomial time***

  (in number and length of the genotypes)

A more difficult problem

- **Imperfect Phylogenetic Tree** :
  some nodes do not appear in the original data
- example : suppose proposed haplotypes are
  00000, **01001**, **11000**, **11010**
  - to explain as a phylogenetic tree, the data needs an extra (”non-observed”) node :



A more difficult problem

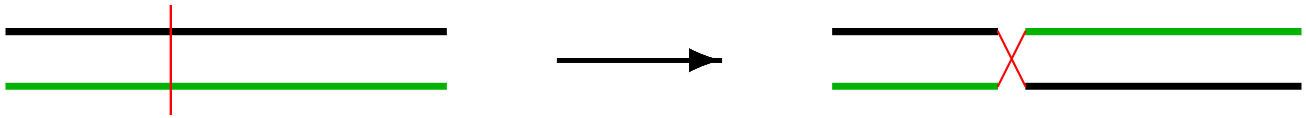
- **Imperfect Phylogenetic Tree** :
  some nodes do not appear in the original data

Theorem (Kimmel & Shamir, 2005)

- *the **Imperfect Phylogeny Haplotype Problem** is **NP-complete***

Further complications

- new haplotypes can occur because of **recombination** :



  - can be modelled with **cycles** in the phylogenetic tree
  - is assumed to happen very rarely

- leads to more and more **combinatorial (decision) problems**

From now on, mathematics only

a simple **ordered colouring** problem

- **given** : graph $G = (V, E)$
  with vertices $V = \{1, 2, \ldots, n\}$

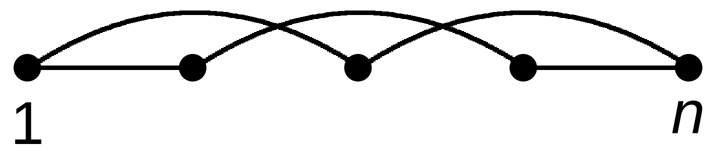
- **required** : a vertex colouring with colours $\{A, B, C\}$
  i.e., a function $\varphi : V \to \{A, B, C\}$

  such that :

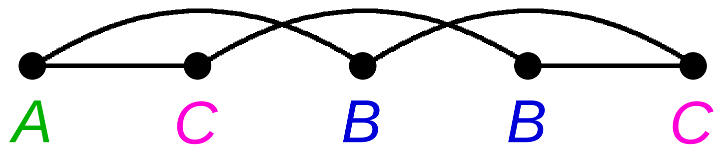
  - $\forall uv \in E : \varphi(u) \neq \varphi(v)$ (i.e., a proper colouring)
  - all vertices coloured A must appear before all vertices coloured B

Example

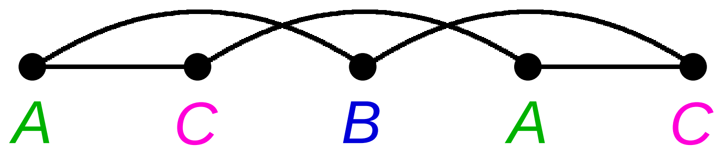
- suppose the graph is :



  - allowed colouring :



  - not allowed :



decision problem :

- given a graph G with vertices $\{1, 2, \ldots, n\}$
  how easy is it to decide if it allows such a 3-colouring ?

Deciding ordered 3-colouring

Theorem

- *deciding if a given graph allows this kind of 3-colouring can be done in* ***polynomial time***

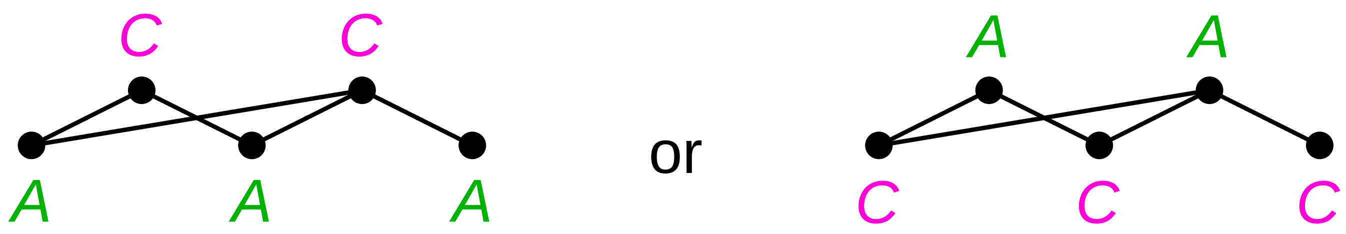
sketch of proof

- choose a vertex m from $1, \ldots, n$ (n choices)
- assume m forms the boundary between
  the vertices that can have colour A ($\leq m$)
  and the vertices that can have colour B ($> m$)

Deciding ordered 3-colouring

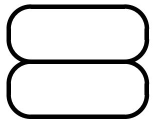
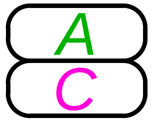
sketch of proof (cont.)

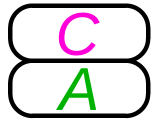
- then vertices $1, \ldots, m$ can only be coloured A or C
  and vertices $m + 1, \ldots, n$ can only be coloured B or C
  - so check if the two parts are **bipartite**
- suppose "**yes**"
  - so $G[1, \ldots, m]$ has bipartite components
    same for $G[m + 1, \ldots, n]$
- each component has 2 different choices for a 2-colouring

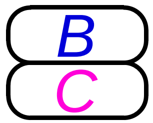


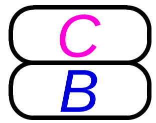
Deciding ordered 3-colouring

sketch of proof (cont.)

- for each component X_i in $G[1, \ldots, m]$
  - choose a **top** and a **bottom** part : 
  - introduce a **Boolean variable** x_i with the meaning
    - $x_i = \text{TRUE}$: 

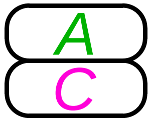
      $x_i = \text{FALSE}$: 

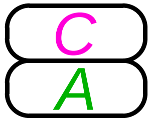
- for each component Y_j in $G[m+1, \ldots, n]$
  - choose a **top** and a **bottom** part
  - introduce a **Boolean variable** y_j with the meaning
    - $y_j = \text{TRUE}$: 

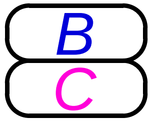
      $y_j = \text{FALSE}$: 

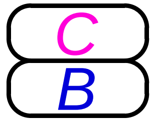
Deciding ordered 3-colouring

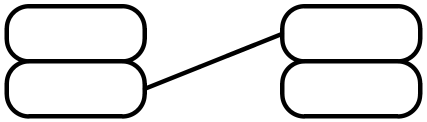
sketch of proof (cont.)

- x_i : 

$\overline{x_i}$: 

y_j : 

$\overline{y_i}$: 

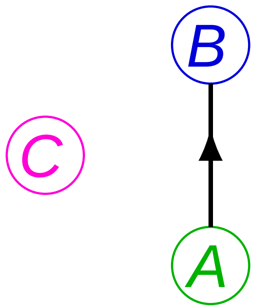
- each edge between a component X_i and a component Y_j gives rise to a **2-literal disjunction**
  - **e.g.** :  $\longleftrightarrow$ $\overline{x_i} \vee y_j$

- doing this for **all edges** between X_i and Y_j components gives a **conjunction of 2-literal disjunctions**

- i.e., an instance of the **2-SATISFIABILITY** problem
  - well-known to be **polynomial time** decidable

Generalising the problem

- a way to look at this colouring problem is as if there is an **order relation** on the colours $\{A, B, C\}$

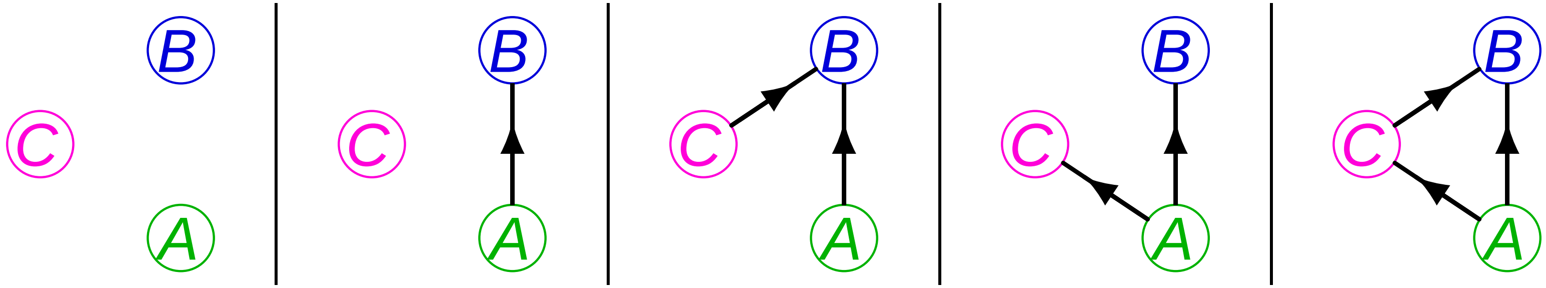


means: "all A before all B"
"A - C and B - C are indifferent"

- what would happen if we consider other order relations on the colours?

More general ordered 3-colourings

- possible posets with 3 elements :



Theorem

- ***deciding*** *if a certain graph with vertices* $1, \ldots, n$ *has a colouring according to these posets is*
  - ***NP-complete*** *for the first poset (just 3-colouring)*
  - ***polynomial*** *for the others*

General ordered colourings with more colours

- **strong order** : $A \longrightarrow B$ means : "all A before all B"
- suppose we allow **any number** of colours and **any poset** on the set of colours

Theorem

- *deciding if a given graph with vertices $1, \ldots, n$ has a colouring according to a **fixed colour poset** is*
  - ***NP-complete***
    if the poset has an anti-chain of length at least 3
  - ***polynomial** otherwise*

A different ordered 3-colouring problem

- **given** : graph $G = (V, E)$ with n vertices
  vertices are numbered from $1, \dots, n$

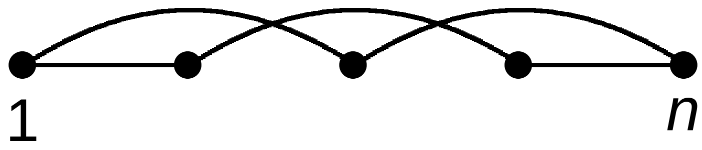
- **required** : a vertex colouring with colours $\{A, B, C\}$
  i.e., a function $\varphi : V \rightarrow \{A, B, C\}$

such that :

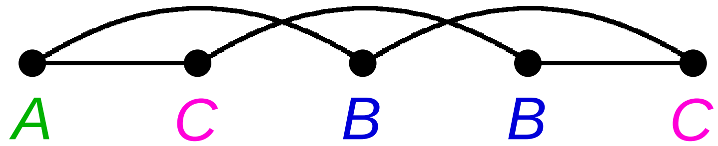
- $\forall uv \in E \ : \ \varphi(u) \neq \varphi(v)$ (i.e., a proper colouring)
- vertices coloured A must appear before vertices coloured B **only for adjacent vertices**

Example

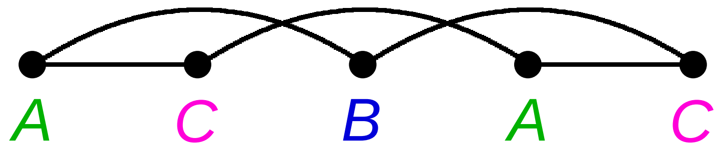
- suspose the graph is :



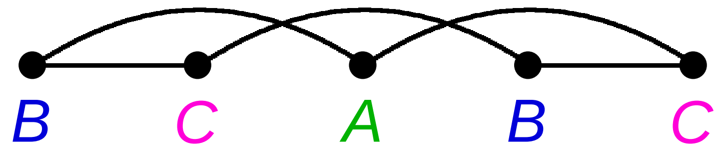
  - allowed colouring :



  - also allowed :



  - not allowed :



A different ordered 3-colouring problem

Theorem

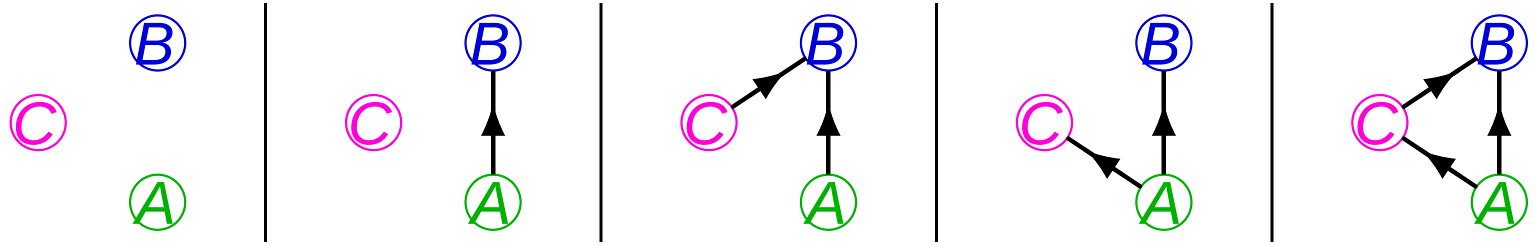
- *this new 3-colouring problem is* ***NP-complete***

ideas of proof

- constraint on edges only means “local constraints”
  - allows to construct gadgets
- reduction to SAT

Generalising the 2nd variant for 3 colours

- **weak order** : $A \longrightarrow B$ means : "A before B **on edges**"
- possible posets with 3 elements :



Theorem

- ***deciding*** *if a certain graph with vertices* $1, \ldots, n$ *has a colouring according to the weak order rule is*
  - ***NP-complete*** *for the first two posets*
  - ***polynomial*** *for the others*

Generalising the 2nd variant with more colours

- **weak order** : $A \longrightarrow B$ means : "A before B **on edges**"
- suppose we allow **any number** of colours and **any poset** on the set of colours

Theorem

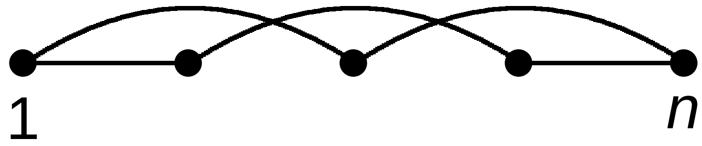
- *deciding if a given graph with vertices $1, \ldots, n$ has a colouring according to a **fixed colour poset** and the weak order rule is*
  - ***polynomial** if*
    *there is **at most one pair of non-comparable colours***
  - ***NP-complete** otherwise*

And now for something completely different . . .

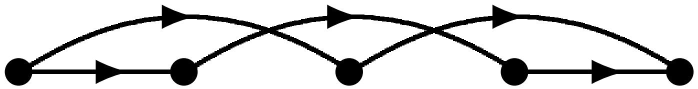
- given two directed graphs $\overrightarrow{G}$ and $\overrightarrow{H}$
- a **directed homomorphism $\overrightarrow{G} \longrightarrow \overrightarrow{H}$** is a function $\psi : V(\overrightarrow{G}) \to V(\overrightarrow{H})$ such that
  - $\overrightarrow{uv}$ an arc in $\overrightarrow{G}$ $\Longrightarrow$ $\overrightarrow{\psi(u)\psi(v)}$ an arc in $\overrightarrow{H}$
- directed homomorphisms seem to be much harder to understand than their undirected cousins

From colourings to homomorphisms

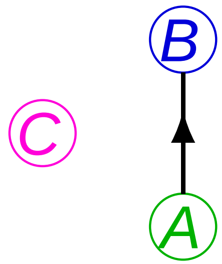
- given graph G



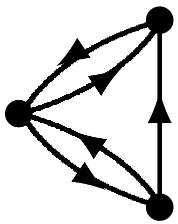
form $\overrightarrow{G}$:



- given colour poset



form $\overrightarrow{H}$:



- **then** :

G is colourable with the poset using the weak order rule

$$\Longleftrightarrow$$

there is a directed homomorphism $\overrightarrow{G} \longrightarrow \overrightarrow{H}$

Some technical definitions

- semi-complete digraph :

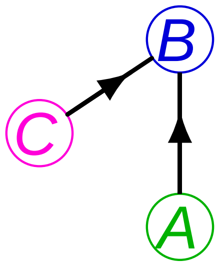
  between any two vertices one arc or two (opposite) arcs

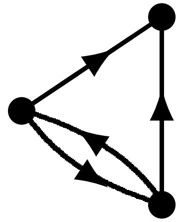
- quasi-acyclic digraph :

  removal of all 2-cycles gives an acyclic digraph

- starting from a colour poset,

  always gives a quasi-acyclic semi-complete digraph $\overrightarrow{H}$

- e.g.: 

  gives $\overrightarrow{H}$: 

Complexity of directed homomorphism

- $\overrightarrow{H}$-COLOURING

Input : directed graph $\overrightarrow{G}$

Question : is there a homomorphism $\overrightarrow{G} \longrightarrow \overrightarrow{H}$?

Theorem (Bang-Jensen, Hell & MacGillivray, 1988)

- *let $\overrightarrow{H}$ be a semi-complete digraph then the complexity of $\overrightarrow{H}$-COLOURING is*
  - ***polynomial** if $\overrightarrow{H}$ contains **at most one directed cycle***
  - ***NP-complete** otherwise*

Complexity of acyclic directed homomorphism

- ACYCLIC-$\overrightarrow{H}$-COLOURING

  Input: acyclic directed graph $\overrightarrow{G}$

  Question: is there a homomorphism $\overrightarrow{G} \longrightarrow \overrightarrow{H}$?

Theorem (corollary of our earlier result)

- *let $\overrightarrow{H}$ be a quasi-acyclic semi-complete digraph then the complexity of ACYCLIC-$\overrightarrow{H}$-COLOURING is*
  - ***polynomial** if $\overrightarrow{H}$ contains **at most one 2-cycle***
  - ***NP-complete** otherwise*

Open problems

- is there a characterisation for semi-complete digraphs $\overrightarrow{H}$ on the complexity of ACYCLIC-$\overrightarrow{H}$-COLOURING ?

  (probably "yes")

- what about a characterisation for general digraphs $\overrightarrow{H}$ on the complexity of ACYCLIC-$\overrightarrow{H}$-COLOURING ?

  (likely to be very hard; open even if we restrict $\overrightarrow{H}$ to trees)

- the Holy Grail :
  characterisation for general digraphs $\overrightarrow{H}$ on the complexity of $\overrightarrow{H}$-COLOURING