

Main page

Contents

Featured content

Current events

Random article

Interaction

Print/export

Languages

Toolbox

Labyrinthulomycetes WIKIPEDIA From Wikipedia, the free encyclopedia The Free Encyclopedia

Wikipedia interlanguage links

The Labyrinthulomycetes (ICBN) or Labyrinthule (ICZN), or Slime nets are a class of protists that pro a network of filaments or tubes,<sup>[2]</sup> which serve as tra for the cells to glide along and absorb nutrients for th There are two main groups, the labyrinthulids and thraustochytrids. They are mostly marine, commonly Donate to Wikipedia found as parasites on alga and seagrass or as decomposers on dead plant material. They also inclu some parasites of marine invertebrates.

> Although they are outside the cells, the filaments are surrounded by a membrane. They are formed and connected with the cytoplasm by a unique organelle a sagenogen or bothrosome. The cells are uninuclea and typically ovoid, and move back and forth along t amorphous network at speeds varying from 5-150 µr minute. Among the labyrinthulids the cells are enclos



### **Observation**

If there is a link path from a word in some language to a different word in the same language, then at least one of the links on this path is wrong: e.g.

Ham – Jambon – Schinken – Prosciutto

### **Optimization problem**

**Colorful Components** 

**Instance:** A network where each node has a color.



Jambon de Parme Пармская ветчина Can we fix them automatically?

**Task:** Delete a minimum number of links such that all connected components are colorful, that is, contain each color only once.

### **Obstacle**

Colorful Components is NP-hard! Thus, there is probably no efficient algorithm that always gives an optimal solution.

### **Integer Linear Programming (ILP)**

**Idea:** Express the problem using linear constraints and a linear objective and use an ILP solver.

# Optimum

### **Parameterized complexity**

**Idea:** Analyze the running time not only with respect to the problem size n, but also with respect to some parameter k, e.g. the number of colors c or the number of link deletions d.

### **Fixed-Parameter Algorithms (FPT)**

**Idea:** Try to confine the combinatorial explosion to some parameter k.



### **FPT algorithm for Colorful Components**

### **Data reduction/Kernelization**

Idea: Remove redundant parts of the input, such that the size of the remaining instance depends only on some parameter k.



## S Ð U 0 Appl

## Methods

ults

S

Re

### **Colorful Components ILP formulation**

Use binary variables  $e_{uv}$  for each node pair (u, v), where  $e_{\mu\nu} = 1 \Leftrightarrow u$  and  $\nu$  are in the same cluster.

maximize  $\sum_{\{u, v\} \in E} e_{uv}$ subject to  $e_{uv} = 0$  when color(u) = color(v) $e_{uv} + e_{uw} - e_{vw} \le 1$  for nodes u, v, w

### Improvements

Use cutting planes and row generation.

### solved (%) instances Implicit Hitting Set litting Set row generation 20 **Clique Partitioning ILP** Clique Partitioning without cuts Branching 10<sup>-2</sup> $10^{-1}$ $10^{0}$ $10^{1}$ $10^{2}$ time (s)

Idea: Find a path between equally-colored nodes (bad path) and recursively try deleting each link.

### Improvement

**Idea:** If there is a node with at least three neighbors, we can find a bad path with at most c-1 links; otherwise, the instance is easy.

### Theorem

Colorful Components can be solved in  $O((c-1)^d \cdot n^2)$  time.

### Implication

We can find optimal solutions with useful running time guarantees: if the number of colors and the number of links deleted are small, we can solve the problem quickly, even if the network is very large.

### **Data reduction for Colorful Components**

**Idea:** Join nodes that cannot be split.



### Theorem

Colorful Components has a kernel with at most *c*·*d* nodes.

### Implication

We have an efficient preprocessing with useful quality guarantees, which can be combined with any other approach, be it exact, approximative or heuristic.

### **Running time for random instances**

### **Final result for the Wikipedia network**

Using a combination of data reduction and the ILP formulation, we can optimally solve in 80 minutes Colorful Components for the Wikipedia interlanguage link network of the 30 most popular languages with 11,977,500 nodes and 46,695,719 links. The largest connected component has 1,828 nodes and 14,403 links. We find 618,660 links to be deleted and 434,849 to be added.

Outlook: Other applications	Outlook: Model extensions	<b>Outlook: Algorithmic improvements</b>
<ul> <li>Matching of products from online store</li> <li>Matching profiles from different social networks</li> <li>Multiple sequence alignment</li> </ul>	<ul> <li>Demand better connected clusters</li> <li>Allow some duplicates per cluster</li> <li>Allow link weights</li> </ul>	<ul> <li>Column generation</li> <li>More data reduction rules</li> <li>Heuristics</li> </ul>

### Reference

Sharon Bruckner, Falk Hüffner, Christian Komusiewicz, Rolf Niedermeier, Sven Thiel, and Johannes Uhlmann: Partitioning into colorful components by minimum edge deletions. In Proceedings of the 23rd Annual Symposium on Combinatorial Pattern Matching (CPM '12), Helsinki, Finland. July 2012. Volume 7354 in Lecture Notes in Computer Science, pages 56–69, Springer, 2012.