Signaling	Pathways
000000	

Algorithm Engineering

Experiments

Algorithm Engineering for Color-Coding to Facilitate Signaling Pathway Detection

Falk Hüffner Sebastian Wernicke Thomas Zichner

Friedrich-Schiller-Universität Jena

Fifth Asia Pacific Bioinformatics Conference January 17, 2007

1/22

イロト 人間ト イヨト イヨト

Signaling	Pathways
000000	

Algorithm Engineering

Experiments

Outline



- Protein Interaction Networks
- Signaling Pathways
- Graph Model

2 Color-Coding

- 3 Algorithm Engineering
 - Worst-case Speedup
 - Lower Bounds

4 Experiments

- Protein Interaction Networks
- Simulations

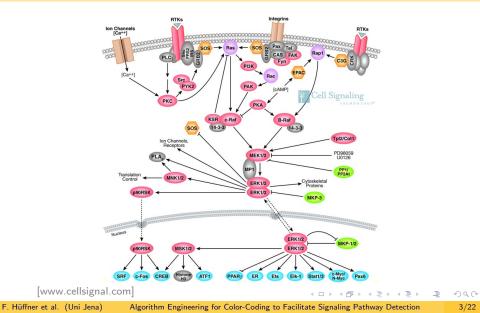
3

通 ト イヨ ト イヨト

Signaling	Pathways
•••••	

Algorithm Engineering

Protein Interaction Networks



3/22

Color-Coding

Algorithm Engineering

Experiments

Protein Interaction Networks

Representation of protein interactions as a graph:

- Proteins are nodes
- Interactions are edges
- Edges are annotated with interaction probability (obtained by two-hybrid screening)

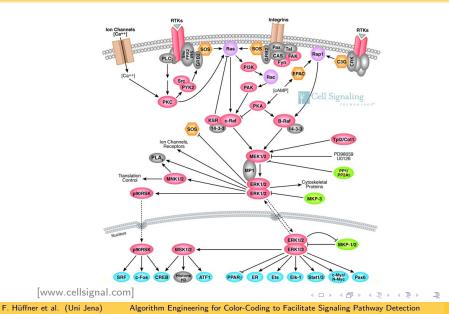
イロト イポト イヨト イヨト

Color-Coding

Algorithm Engineering

Experiments

Signaling Pathways



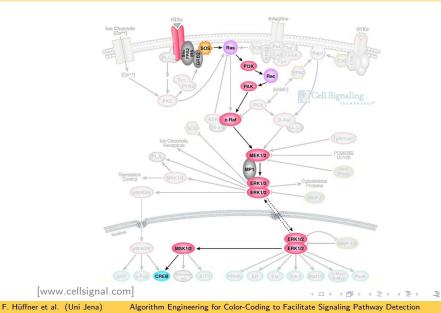
5/22

Color-Coding

Algorithm Engineering

Experiments 00000

Signaling Pathways



∽へで 5/22

Color-Coding

Algorithm Engineering

・ロト ・ 一日 ト ・ 日 ト ・ 日 ・

Experiments

6/22

Signaling Pathways

Sequence of distinct proteins, where each interacts strongly with the previous one.

Most Probable Path

Input: Graph G = (V, E), interaction probabilities $p : E \to [0, 1]$, integer k > 0. **Task:** Find a non-overlapping path v_1, \ldots, v_k of length k in G that maximizes $p(v_1, v_2) \cdot \ldots \cdot p(v_{k-1}, v_k)$.

Color-Coding

Algorithm Engineering

Experiments

Signaling Pathways

Sequence of distinct proteins, where each interacts strongly with the previous one.

Most Probable Path

Input: Graph G = (V, E), interaction probabilities $p : E \rightarrow [0, 1]$, integer k > 0.

Task: Find a non-overlapping path v_1, \ldots, v_k of length k in G that maximizes $p(v_1, v_2) \cdot \ldots \cdot p(v_{k-1}, v_k)$.

Setting $w(e) := -\log(p(e))$:

MINIMUM-WEIGHT PATH

Input: Graph G = (V, E), weights $w : E \to [0, 1]$, integer k > 0. Task: Find a non-overlapping path v_1, \ldots, v_k of length k in G that minimizes $w(v_1, v_2) + \cdots + w(v_{k-1}, v_k)$.

3

イロト イポト イヨト イヨト

Signaling	Pathways
000000	

Yeast Network

Color-Coding

Algorithm Engineering

Experiments

 $4\,400$ proteins, $14\,300$ interactions, looking for paths of length 5--15

Color-Coding

Algorithm Engineering

(4 回) (4 日) (4 日)

Experiments 00000

8/22

Minimum-Weight Path

Theorem

MINIMUM-WEIGHT PATH is NP-hard [Garey&Johnson 1979].

For an exact algorithm, we have to accept exponential runtime.

Idea

Exploit the fact that the paths sought for are rather short (\approx 5–15): restrict the exponential part of the runtime to k (parameterized complexity).

イロト 人間ト イヨト イヨト

9/22

Color-Coding

Color-coding [Alon, Yuster&Zwick J. ACM 1995]:

- randomly color each vertex of the graph with one of k colors
- hope that all vertices in the subgraph searched for obtain different colors (colorful)
- solve the MINIMUM-WEIGHT PATH under this assumption (which is much quicker)
- repeat until it is reasonably certain that the path was colorful at least once

Result: exponential part of the runtime depends only on k

Signaling Pathways	Color-Coding ⊙●○○	Algorithm Engineering	Experiments 00000
Dynamic Progr	amming for M	inimum-Weight Co	lorful Path

Idea

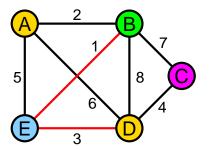
Table entry W[v, C] stores the minimum-weight path that ends in v and uses exactly the colors in S.

• • = • • = •

Signaling Pathways	Color-Coding	Algorithm Engineering	Experiments
	○●○○	00000	00000
Dynamic Program	nming for Mini	imum-Weight Color	ful Path

Idea

Table entry W[v, C] stores the minimum-weight path that ends in v and uses exactly the colors in S.



$$W[B, \{ \bigcirc, \bigcirc, \bigcirc \}] = 4$$

Signaling Pathways	Color-Coding	Algorithm Engineering	Experiments 00000
Dynamic Prograu	mming for Min	imum-Weight Colo	rful Path

Coloring
$$c: V \to \{1, \ldots, k\}$$

Recurrence

 $W[v, C] = \min_{u \in N(v) | c(u) \in C \setminus \{c(v)\}} (W[u, C \setminus \{c(v)\}] + w(u, v))$

- ◆ 母 ▶ ◆ 臣 ▶ ◆ 臣 ▶ ○ 臣 = ∽ � � �

Signaling Pathways	Color-Coding	Algorithm Engineering	Experiments 00000
Dynamic Program	nming for Mini	imum-Weight Color	ful Path

Coloring
$$c: V \to \{1, \ldots, k\}$$

Recurrence

$$W[v, C] = \min_{u \in N(v) | c(u) \in C \setminus \{c(v)\}} (W[u, C \setminus \{c(v)\}] + w(u, v))$$

- Each table entry can be calculated in O(n) time
- $n2^k$ table entries

$$\rightsquigarrow$$
 Runtime: $O(n \cdot n2^k) = n^2 \cdot 2^k$

(1日) (1日) (1日)

3

Signaling Pathways 000000	Color-Coding	Algorithm Engineering	Experi 0000
Color-coding R	untime		

- $O(n^2 \cdot 2^k)$ time per trial
- To obtain error probability ε , one needs $O(|\ln \varepsilon| \cdot e^k)$ trials

Theorem ([ALON et al. JACM 1995])

MINIMUM-WEIGHT PATH can be solved in $O(|\ln \varepsilon| \cdot 5.44^k |G|)$ time).

通 ト イヨ ト イヨト

Signaling	Pathways
000000	

Algorithm Engineering

Experiments

- 4 週 ト - 4 ヨ ト - 4 ヨ ト - -

12/22

Color-coding Runtime

- $O(n^2 \cdot 2^k)$ time per trial
- To obtain error probability ε , one needs $O(|\ln \varepsilon| \cdot e^k)$ trials

Theorem ([ALON et al. JACM 1995])

MINIMUM-WEIGHT PATH can be solved in $O(|\ln \varepsilon| \cdot 5.44^k |G|)$ time).

Color-coding can find minimum-weight paths of length 10 in the yeast protein interaction networks within 3 hours $(n = 4\,400, k = 10)$ [Scott et al., RECOMB'05]

Color-Coding

Algorithm Engineering

イロト イポト イヨト イヨト

-

Experiments

Increasing the Number of Colors

Idea

Use k + x colors instead of k colors.

Trial runtime:

$$O(2^k|G|) \rightarrow O(2^{k+x}|G|)$$

Color-Coding

Algorithm Engineering

Experiments

Increasing the Number of Colors

Idea

Use k + x colors instead of k colors.

Trial runtime:

$$O(2^k|G|) \rightarrow O(2^{k+x}|G|)$$

Probability P_c for colorful path (k = 8, $\varepsilon = 0.001$):

x	0	1	2	3	4	5
P _c	0.0024	0.0084	0.0181	0.0310	0.0464	0.0636
trials	2871	816	378	220	146	106

13/22

3

イロト イポト イヨト イヨト

Color-Coding

Algorithm Engineering

Experiments

Increasing the Number of Colors

Idea

Use k + x colors instead of k colors.

Trial runtime:

$$O(2^k|G|) \rightarrow O(2^{k+x}|G|)$$

Probability P_c for colorful path (k = 8, $\varepsilon = 0.001$):

x	0	1	2	3	4	5
P _c	0.0024	0.0084	0.0181	0.0310	0.0464	0.0636
trials	2871	816	378	220	146	106

Theorem

MINIMUM-WEIGHT PATH can be solved in $O(|\ln \varepsilon| \cdot 4.32^k |G|)$ time by choosing x = 0.3k.

(4 同) トイヨト イヨト

Color-Coding

Algorithm Engineering

A (10) F (10)

Experiments

Increasing the Number of Colors

Idea

Use k + x colors instead of k colors.

Trial runtime:

$$O(2^k|G|) \rightarrow O(2^{k+x}|G|)$$

Probability P_c for colorful path (k = 8, $\varepsilon = 0.001$):

x	0	1	2	3	4	5
P _c	0.0024	0.0084	0.0181	0.0310	0.0464	0.0636
trials	2871	816	378	220	146	106

Theorem

MINIMUM-WEIGHT PATH can be solved in $O(|\ln \varepsilon| \cdot 4.32^k |G|)$ time by choosing x = 0.3k.

But: Higher memory usage

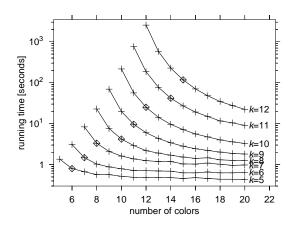
F. Hüffner et al. (Uni Jena) Algorithm Engineering for Color-Coding to Facilitate Signaling Pathway Detection 13/22

Color-Coding

Algorithm Engineering

Experiments 00000

Increasing the Number of Colors



Runtimes for the yeast protein interaction network (highlighted point of each curve marks worst-case optimum)

14/22

Signaling Pathways	Color-Coding
000000	0000

Algorithm Engineering

イロト イポト イヨト イヨト

-

Exploiting Lower Bounds

Idea

Use a known solution to prune "hopeless" table entries.

• Discard entries that already have a weight higher than the known solution.

Color-Coding

Algorithm Engineering

イロト イポト イヨト イヨト

Experiments

Exploiting Lower Bounds

Idea

Use a known solution to prune "hopeless" table entries.

- Discard entries that already have a weight higher than the known solution.
- Discard entries when

weight + (minimum edge weight · edges left)

is higher than the weight of the known solution.

Color-Codin

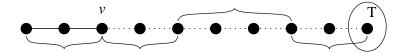
Algorithm Engineering

・ 同 ト ・ ヨ ト ・ ヨ ト

Experiments

Precalculated Lower Bounds

For each vertex u and a range of lengths $1 \le i \le d$, determine the minimum weight of a path of i edges that starts at u.



Color-Codin

Algorithm Engineering

< 17 ▶

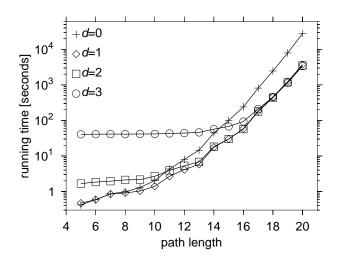
3

1

3

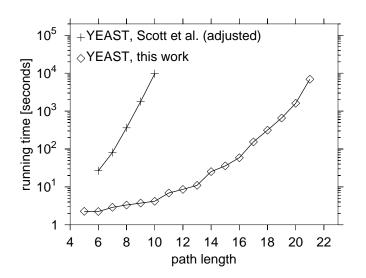
Experiments

Lower Bounds Experiments



F. Hüffner et al. (Uni Jena) Algorithm Engineering for Color-Coding to Facilitate Signaling Pathway Detection

Signaling Pathways	Color-Coding	Algorithm Engineering
000000	0000	00000
Yeast Network		



3

Experiments

Signaling Pathways	Color-Coding	Algorithm Engineering	Experiments	
000000	0000	00000	00000	
Network Comparis	son			

	V	E	clust. coeff.	avg. degree	max. degree
	4 389	14 319	0.067	6.5	237
-	7 009	20 440	0.030	5.8	175

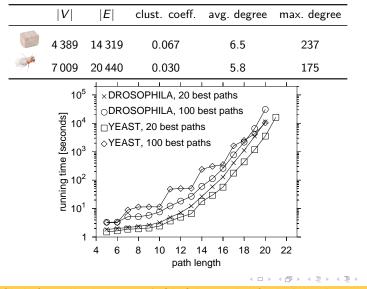
▲□▶ ▲□▶ ▲目▶ ▲目▶ ▲目▶ ▲□

Color-Codin

Algorithm Engineering

Experiments

Network Comparison

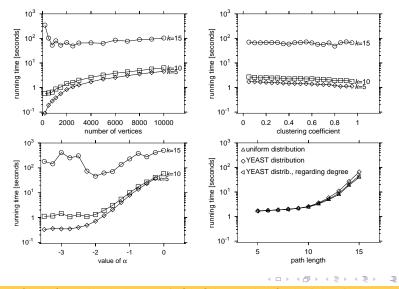


F. Hüffner et al. (Uni Jena) Algorithm Engineering for Color-Coding to Facilitate Signaling Pathway Detection

19/22

Circulations	Dobustness of A	Les e stitle see	
Signaling Pathways	Color-Coding	Algorithm Engineering 00000	Experiments

Simulations: Robustness of Algorithm



F. Hüffner et al. (Uni Jena) Algorithm Engineering for Color-Coding to Facilitate Signaling Pathway Detection

Color-Coding

Algorithm Engineering

イロト イポト イヨト イヨト

3

Experiments

Conclusion & Outlook

Color-coding, with some algorithm engineering, is a practical and reliable method for finding signaling pathways in protein interaction networks.

Color-Codin

Algorithm Engineering

イロト イポト イヨト イヨト

Experiments

Conclusion & Outlook

Color-coding, with some algorithm engineering, is a practical and reliable method for finding signaling pathways in protein interaction networks.

Future work:

- Pathway queries
- Richer motifs (cycles, trees, ...)
- Derandomization

Color-Coding

Algorithm Engineering

Experiments

Graphical User Interface (upcoming)

				Fast Signal	ing Pathw	ay Detectio	n			_ 0
le	∑iew Help									
	H	• •	127	72.3	1 -					
Opt	tions Inform	ation		Graph	1 Graph 2	Graph 3	Graph 4 G	raph 5 Gra	ph 6 Graph	7
Ma	ain Start no bad Graph iome/tzsnoop Pathle Number of p	des End no py/uni/reposi angth 8 paths 50 Filter 70	tory/colorcod							
_			el tab	Ŧ		1112				,
	Weight	Prot 1	Prot 2	Prot 3	Prot 4	Prot 5	Prot 6	Prot 7	Prot 8	
-								CG11761		Selected
	0.317429	CG6998	CG3227	CG5450	CG32130				CG5053	Selected
-		CG6998 CG1871	CG3227 CG8929	CG5450 CG13030	CG32130 CG10108	CG18743 C CG1856	CG7057	CG13811	CG5063 CG3779	
2	0.317429 0.323947	CG1871	CG8929	CG13030	CG10108	CG1856			CG3779	□ ☑
2	0.317429			CG13030 CG7945	CG10108 CG11761		CG7057	CG13811	CG3779 CG11454	
2	0.317429 0.323947 0.339116 0.368402	CG1871 CG32130 CG5450	CG8929 CG18743[C CG32130	CG13030 CG7945 CG18743 C0	CG10108 CG11761 CG7945	CG1856 CG17599 CG11761	CG7057 CG9740 CG1435	CG13811 CG4622 CG2774	CG3779 CG11454 CG8282	2 2 2
2 3	0.317429 0.323947 0.339116	CG1871 CG32130 CG5450 CG15283	CG8929 CG18743 C	CG13030 CG7945 CG18743 C0	CG10108 CG11761	CG1856 CG17599	CG7057 CG9740 CG1435 CG7057	CG13811 CG4622	CG3779 CG11454 CG8282 CG3779	
2 3 4 5 6	0.317429 0.323947 0.339116 0.368402 0.373786	CG1871 CG32130 CG5450	CG8929 CG18743[C CG32130 CG14168	CG13030 CG7945 CG18743[C0 CG7224	CG10108 CG11761 CG7945 CG13030	CG1856 CG17599 CG11761 CG1856	CG7057 CG9740 CG1435	CG13811 CG4622 CG2774 CG13811	CG3779 CG11454 CG8282	
1 2 3 4 5 6 7 8	0.317429 0.323947 0.339116 0.368402 0.373786 0.391802	CG1871 CG32130 CG5450 CG15283 CG15468	CG8929 CG18743[C CG32130 CG14168 CG14818	CG13030 CG7945 CG18743[C0 CG7224 CG9951	CG10108 CG11761 CG7945 CG13030 CG6856 CG6610	CG1856 CG17599 CG11761 CG1856 CG17599	CG7057 CG9740 CG1435 CG7057 CG9740	CG13811 CG4622 CG2774 CG13811 CG4622	CG3779 CG11454 CG8282 CG3779 CG11454	

F. Hüffner et al. (Uni Jena)

・ロト ・ 日 ト ・ ヨ ト ・ ヨ ト ・

3