Optimal Edge Deletions for Signed Graph Balancing

Falk Hüffner Nadja Betzler Rolf Niedermeier

Friedrich-Schiller-Universität Jena
Institut für Informatik

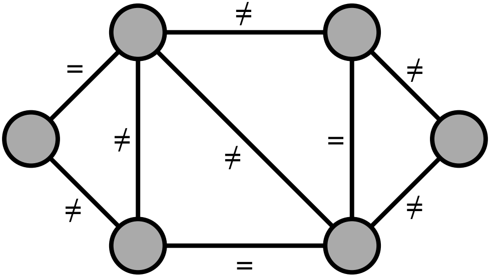
WEA 2007
6th Workshop on Experimental Algorithms
8 June 2007

Outline

Balanced graphs

Definition

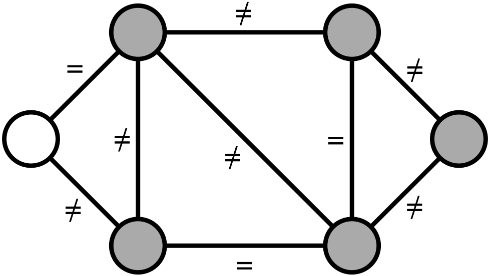
A graph with edges labeled by $=$ or $\neq$ (signed graph) is balanced if the vertices can be colored with two colors such that the relation on each edge holds.



Balanced graphs

Definition

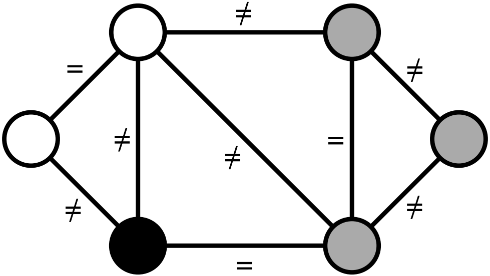
A graph with edges labeled by $=$ or $\neq$ (signed graph) is balanced if the vertices can be colored with two colors such that the relation on each edge holds.



Balanced graphs

Definition

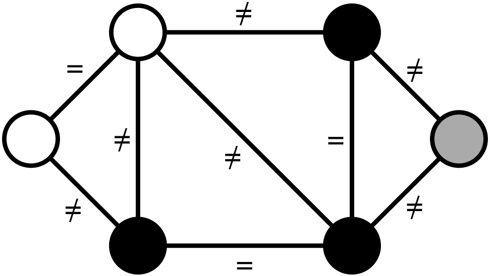
A graph with edges labeled by $=$ or $\neq$ (signed graph) is balanced if the vertices can be colored with two colors such that the relation on each edge holds.



Balanced graphs

Definition

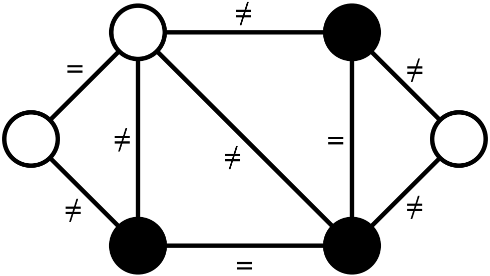
A graph with edges labeled by $=$ or $\neq$ (signed graph) is balanced if the vertices can be colored with two colors such that the relation on each edge holds.



Balanced graphs

Definition

A graph with edges labeled by $=$ or $\neq$ (signed graph) is balanced if the vertices can be colored with two colors such that the relation on each edge holds.



Characterization of balance

Special case

Bipartite graphs are balanced graphs that contain only $\neq$-edges.

Characterization of balance

Special case

Bipartite graphs are balanced graphs that contain only $\neq$-edges.

Theorem (Kőnig 1936)

A signed graph is balanced iff it contains no cycle with an odd number of $\neq$-edges.

Characterization of balance

Special case

Bipartite graphs are balanced graphs that contain only $\neq$-edges.

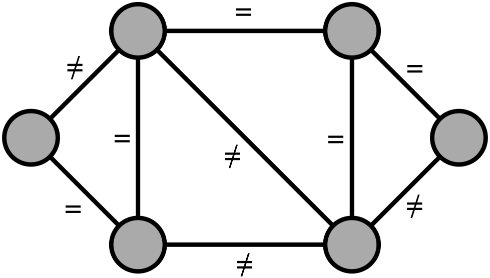
Theorem (Kőnig 1936)

A signed graph is balanced iff it contains no cycle with an odd number of $\neq$-edges.

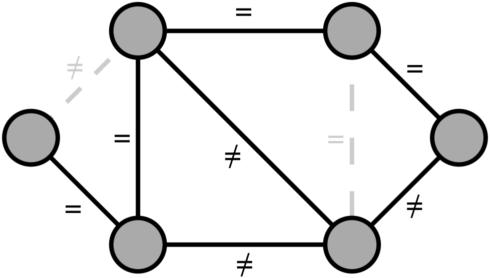
Corollary

Bipartite graphs are graphs that contain no cycle of odd length.

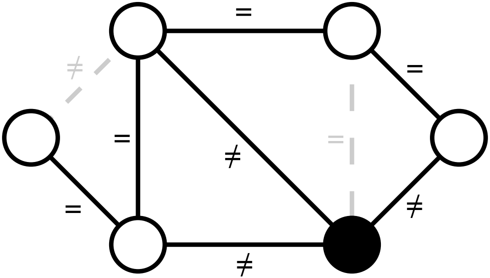
Balanced Subgraph



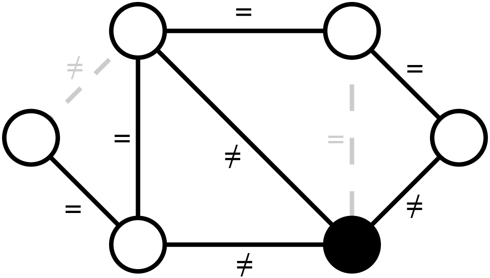
Balanced Subgraph



Balanced Subgraph



Balanced Subgraph



Definition (Balanced Subgraph)

Input: A graph with edges labeled by $=$ or $\neq$.
Task: Find a minimum set of edges to delete such that the graph becomes balanced.

Applications of Balanced Subgraph

- "Monotone subsystems" in biological networks
  [DasGupta et al., WEA 2006]
- Balance in social networks
  [Harary, Mich. Math. J. 1953]
- Portfolio risk analysis
  [Harary et al., IMA J. Manag. Math. 2002]
- Minimum energy state of magnetic materials (spin glasses)
  [Kasteleyn, J. Math. Phys. 1963]
- Stability of fullerenes
  [Došlić&Vikičević, Discr. Appl. Math. 2007]
- Integrated circuit design
  [Chiang et al., IEEE Trans. CAD of IC&Sys. 2007]

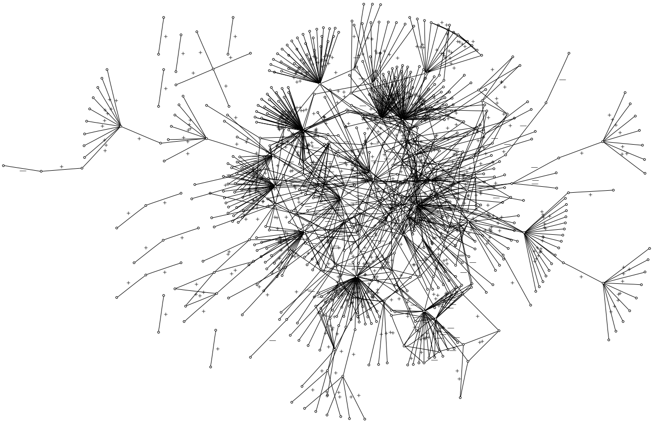
Balanced Subgraph: known results

- BALANCED SUBGRAPH is NP-hard, since it is a generalization of MAX-CUT (MAX-CUT is the special case where all edges are $\neq$)
- A solution that keeps at least 87.8 % of the edges can be found in polynomial time [DASGUPTA et al., WEA 2006]
- A solution that deletes at most c times the edges that need to be deleted can probably not be found in polynomial time [KHOT, STOC 2002]

Graph structure

Idea

Exploit the structure of the relevant networks



Data reduction

Data reduction

Replace the instance by a simpler, equivalent one.

Data reduction

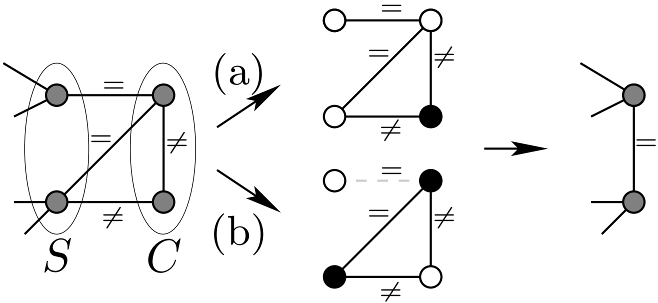
Data reduction

Replace the instance by a simpler, equivalent one.

Example

Delete all degree-1 vertices.

Separator-based data reduction



Data reduction scheme

Data reduction scheme

- Find separator S that cuts off small component C
- For each of the (up to symmetry) $2^{|S|-1}$ colorings of S, determine the size of an optimal solution for $G[S \cup C]$
- Replace in G the subgraph $G[S \cup C]$ by an equivalent smaller gadget

Data reduction scheme

Data reduction scheme

- Find separator S that cuts off small component C
- For each of the (up to symmetry) $2^{|S|-1}$ colorings of S, determine the size of an optimal solution for $G[S \cup C]$
- Replace in G the subgraph $G[S \cup C]$ by an equivalent smaller gadget

Subsumes all 8 data reduction rules given by [Wernicke, 2003] for Edge Bipartization

Filling in the data reduction scheme

- How to find good S/C-combinations?

Filling in the data reduction scheme

- How to find good S/C-combinations?
  - Separators of size 0 and 1 can be found in linear time by depth-first search [Gabow, IPL 2000]
  - Separators of constant size can be found in polynomial time [Henzinger et al., J. Alg. 2000]; however, we use a heuristic for separators up to size 4

Filling in the data reduction scheme

- How to find good S/C-combinations?
  - Separators of size 0 and 1 can be found in linear time by depth-first search [Gabow, IPL 2000]
  - Separators of constant size can be found in polynomial time [Henzinger et al., J. Alg. 2000]; however, we use a heuristic for separators up to size 4
- How to construct gadgets that behave equivalently to $S \cup C$?

Gadget construction

Idea

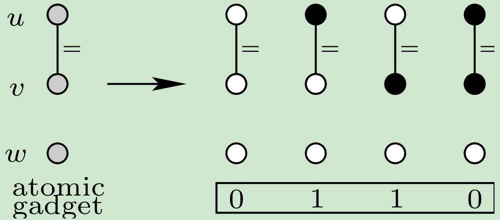
Use atomic gadgets and describe their effect by cost vectors.

Gadget construction

Idea

Use atomic gadgets and describe their effect by cost vectors.

Example

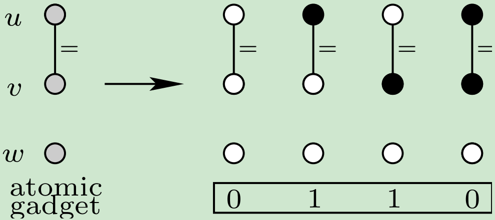


Gadget construction

Idea

Use atomic gadgets and describe their effect by cost vectors.

Example

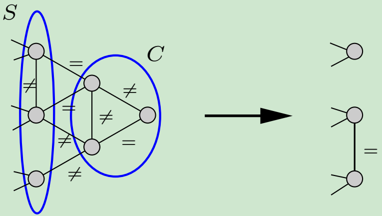


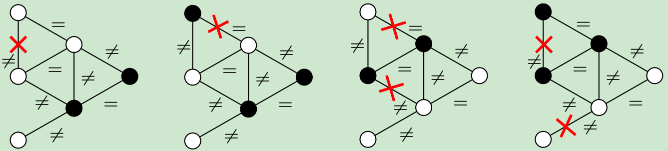
Theorem

With 10 atomic gadgets, we can emulate the behavior of any component behind a 3-vertex cut.

Gadget construction

Example





1	1	2	2	−(1,1,1,1)
0	0	1	1	−(0,0,1,1)
0	0	0	0	

Gadget construction

Problem

How to determine an appropriate set of atomic cost vectors for a given cost vector?

Gadget construction

Problem

How to determine an appropriate set of atomic cost vectors for a given cost vector?

Vector Sum Problem

Given a set S of n vectors of length l with nonnegative integer components and a target vector t of length l, find a sub-(multi)-set of vectors from S that sums to t.

Gadget construction

Problem

How to determine an appropriate set of atomic cost vectors for a given cost vector?

Vector Sum Problem

Given a set S of n vectors of length l with nonnegative integer components and a target vector t of length l, find a sub-(multi)-set of vectors from S that sums to t.

- "Equality-constrained multidimensional knapsack"

Gadget construction

Problem

How to determine an appropriate set of atomic cost vectors for a given cost vector?

Vector Sum Problem

Given a set S of n vectors of length l with nonnegative integer components and a target vector t of length l, find a sub-(multi)-set of vectors from S that sums to t.

- "Equality-constrained multidimensional knapsack"
- In our implementation: simple branch & bound

Gadget construction

Problem

How to determine an appropriate set of atomic cost vectors for a given cost vector?

Vector Sum Problem

Given a set S of n vectors of length l with nonnegative integer components and a target vector t of length l, find a sub-(multi)-set of vectors from S that sums to t.

- "Equality-constrained multidimensional knapsack"
- In our implementation: simple branch & bound
- Sometimes this is a bottleneck!

Gadget construction

Theorem

All separators with $|S| = 2$ and $|C| \geq 1$ and and all separators with $|S| = 3$ and $|C| \geq 2$ are subject to data reduction.

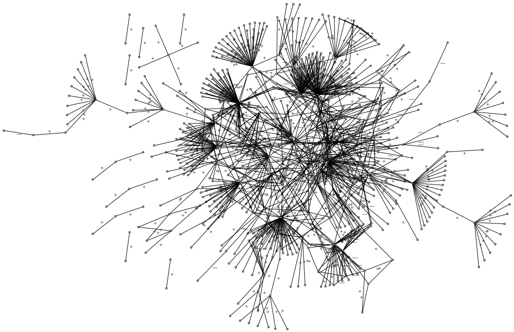
Gadget construction

Theorem

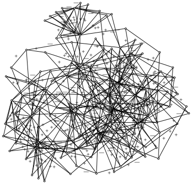
All separators with $|S| = 2$ and $|C| \geq 1$ and and all separators with $|S| = 3$ and $|C| \geq 2$ are subject to data reduction.

- 4-cuts: 2948 atomic gadgets (heuristically found)

Reduction. . . and then?

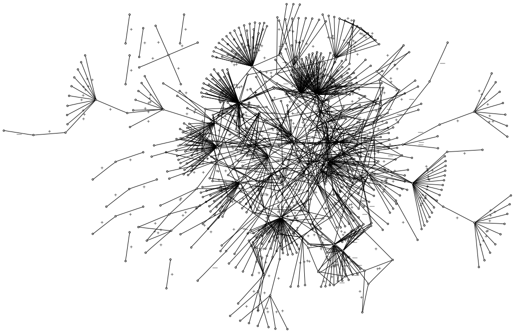


$n = 690,\ m = 1082$

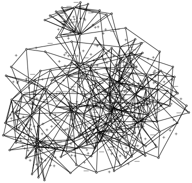


$n = 144,\ m = 405$

Reduction. . . and then?



$n = 690,\ m = 1082$



$n = 144,\ m = 405$

After data reduction, a hard "core" remains.

Fixed-parameter tractability

Idea

Exploit the fact that biological networks are close to being balanced (i. e., the number k of edges that need to be deleted to make them balanced is small).

Fixed-parameter tractability

Parameterized complexity is an approach to finding exact solutions to NP-hard problems by confining the combinatorial explosion to a parameter.

Fixed-parameter tractability

Parameterized complexity is an approach to finding exact solutions to NP-hard problems by confining the combinatorial explosion to a parameter.

Definition ([Downey&Fellows 1999])

A problem is called fixed-parameter tractable with respect to a parameter k if an instance of size n can be solved in $f(k) \cdot n^{O(1)}$ time for an arbitrary function f.

Fixed-parameter tractability

Parameterized complexity is an approach to finding exact solutions to NP-hard problems by confining the combinatorial explosion to a parameter.

Definition ([Downey&Fellows 1999])

A problem is called fixed-parameter tractable with respect to a parameter k if an instance of size n can be solved in $f(k) \cdot n^{O(1)}$ time for an arbitrary function f.

Theorem

BALANCED SUBGRAPH *can be solved in* $O(2^k \cdot m^2)$ *time by a reduction to* EDGE BIPARTIZATION *and using an algorithm based on iterative compression* [Guo et al., JCSS 2006].

Fixed-parameter tractability

Parameterized complexity is an approach to finding exact solutions to NP-hard problems by confining the combinatorial explosion to a parameter.

Definition ([Downey&Fellows 1999])

A problem is called fixed-parameter tractable with respect to a parameter k if an instance of size n can be solved in $f(k) \cdot n^{O(1)}$ time for an arbitrary function f.

Theorem

Balanced Subgraph *can be solved in* $O(2^k \cdot m^2)$ *time by a reduction to* Edge Bipartization *and using an algorithm based on iterative compression* [Guo et al., JCSS 2006].

A heuristic speedup trick can give large speedups over this worst-case running time.

Experimental results

			Approximation			Exact alg.	
Data set	n	m	$k \geq$	$k \leq$	t [min]	k	t [min]
EGFR	330	855	196	219	7	210	108
Yeast	690	1082	0	43	77	41	1
Macr.	678	1582	218	383	44	374	1

Experimental results

			Approximation			Exact alg.	
Data set	n	m	$k \geq$	$k \leq$	t [min]	k	t [min]
EGFR	330	855	196	219	7	210	108
Yeast	690	1082	0	43	77	41	1
Macr.	678	1582	218	383	44	374	1

- Yeast is not solvable without reducing 4-cuts
- A real-world network with 688 vertices and 2208 edges could not be solved

Outlook

- Directed case of Balanced Subgraph
  - Problem: Characterization by two-coloring holds only for strongly connected graphs
- The data reduction scheme is applicable to all graph problems where a coloring or a subset of the vertices is sought. For example:
  - Vertex Cover
  - Dominating Set
  - 3-Coloring
  - Feedback Vertex Set