Algorithmic and Hardness Results for the Colorful Components Problems

Guillaume Blin
Univ. of Bordeaux / CNRS, Talence, France

Anna Adamaszek
University of Copenhagen, Denmark

Alex Popa
Nazarbayev University, Astana, Kazakhstan

TU Wien
July 2015

Problem(s) Definition(s)

Related work

Our Results

Minimum Singleton Vertices

Approximation algorithm for MEC

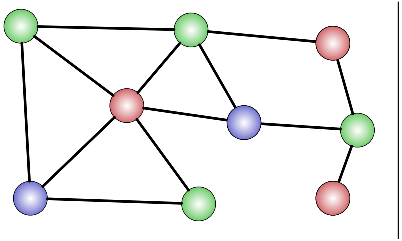
$|V|^{1/3-\varepsilon}$ hardness of approximation for MEC

Hardness of approximation for MEC for 3 colors

Open problems

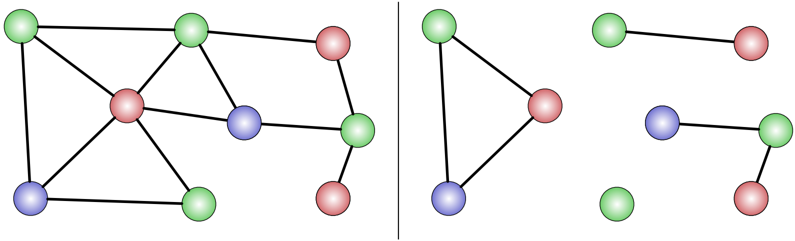
Colorful Components framework

Input: An undirected graph G with colored vertices.
Goal: Delete edges to get *colorful components*.



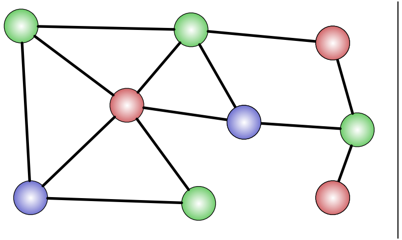
Colorful Components framework

Input: An undirected graph G with colored vertices.
Goal: Delete edges to get *colorful components*.



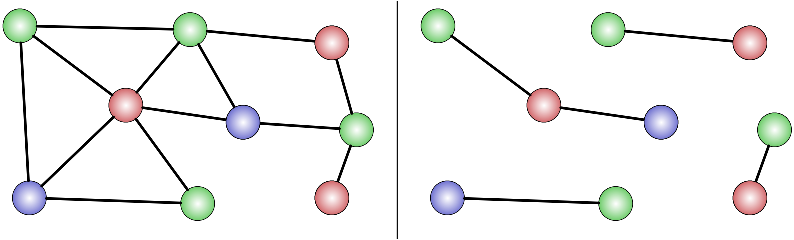
Minimum Singleton Vertices

Goal: Minimize the number of singleton vertices.



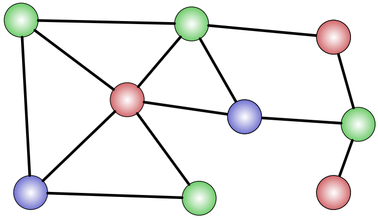
Minimum Singleton Vertices

Goal: Minimize the number of singleton vertices.



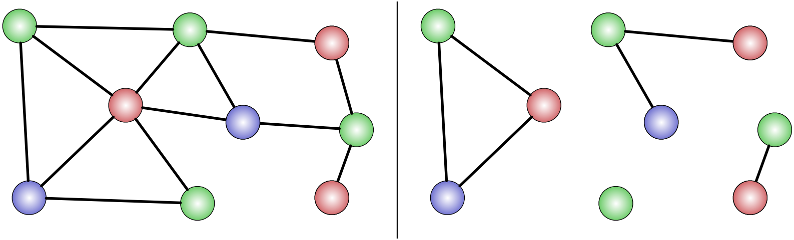
Maximum Edges in Transitive Closure

Goal: Maximize the number of edges in the transitive closure.



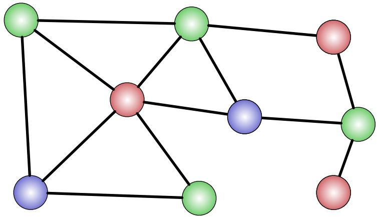
Maximum Edges in Transitive Closure

Goal: Maximize the number of edges in the transitive closure.



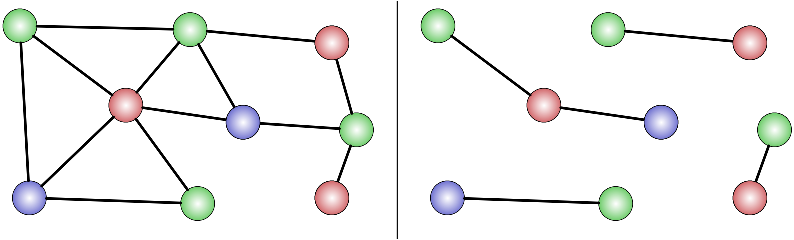
Minimum Colorful Components

Goal: Minimize the number of colorful components.



Minimum Colorful Components

Goal: Minimize the number of colorful components.



Related Work

Colorful Components asks for the minimum number of edge deletions.

- NP-hard for ≥ 3 colors + FPT algorithms [Bruckner, Hüffner, Komusiewicz, Niedermeier, Thiel, Uhlmann 2012]
- APX-hard
- Heuristics [Bruckner, Hüffner, Komusiewicz, Niedermeier, Thiel, Uhlmann 2012] and [Zheng, Swenson, Lyons, Sankoff 2011]
- Approximation algorithms for a special case [He, Liu, Zha 2000]
- $O(\log |C|)$ approximation (from Multi-Multiway Cut)

Related Work

- MSV and MEC are introduced by [Zheng, Swenson, Lyons, Sankoff 2011]
- Applications in comparative genomics
- Heuristics
- MSV and MEC are conjectured to be NP-hard
- 2/3 approximation for the dual of MSV (MAX-OREC) [Tremblay-Savard, Swenson 2012]

Our Results

1. $O(|V| \cdot |E|)$ *exact* algorithm for MSV
   - The conjecture of [Zheng, Swenson, Lyons, Sankoff 2011] is disproved
   - Exact algorithm for MAX-OREC

Our Results

1 $O(|V| \cdot |E|)$ *exact* algorithm for MSV
   - The conjecture of [Zheng, Swenson, Lyons, Sankoff 2011] is disproved
   - Exact algorithm for MAX-OREC

2a MEC is APX is hard for 3 colors
   - Reduction from Maximum Bounded 3-Dimensional Matching
   - Settles the complexity for constant number of colors

Our Results

1. $O(|V| \cdot |E|)$ *exact* algorithm for MSV
   - The conjecture of [Zheng, Swenson, Lyons, Sankoff 2011] is disproved
   - Exact algorithm for MAX-OREC

2a MEC is APX is hard for 3 colors
   - Reduction from Maximum Bounded 3-Dimensional Matching
   - Settles the complexity for constant number of colors

2b MEC is NP-hard to approximate within $|V|^{1/3-\varepsilon}$
   - Holds for trees and even if the color appears at most twice
   - Reduction from the Independent Set

Our Results

1 $O(|V| \cdot |E|)$ *exact* algorithm for MSV
  - The conjecture of [Zheng, Swenson, Lyons, Sankoff 2011] is disproved
  - Exact algorithm for MAX-OREC

2a MEC is APX is hard for 3 colors
  - Reduction from Maximum Bounded 3-Dimensional Matching
  - Settles the complexity for constant number of colors

2b MEC is NP-hard to approximate within $|V|^{1/3-\varepsilon}$
  - Holds for trees and even if the color appears at most twice
  - Reduction from the Independent Set

2c $\sqrt{2 \cdot OPT}$-approximation algorithm for MEC
  - Based on the exact algorithm for MSV

Our Results

1 $O(|V| \cdot |E|)$ *exact* algorithm for MSV
- The conjecture of [Zheng, Swenson, Lyons, Sankoff 2011] is disproved
- Exact algorithm for MAX-OREC

2a MEC is APX is hard for 3 colors
- Reduction from Maximum Bounded 3-Dimensional Matching
- Settles the complexity for constant number of colors

2b MEC is NP-hard to approximate within $|V|^{1/3-\varepsilon}$
- Holds for trees and even if the color appears at most twice
- Reduction from the Independent Set

2c $\sqrt{2 \cdot OPT}$-approximation algorithm for MEC
- Based on the exact algorithm for MSV

3 MCC is hard to approximate within $|V|^{1/2-\epsilon}$
- Reduction from Minimum Clique Partition

A lower bound for MSV

Lemma

In any feasible solution, for any color c, the number of singletons is at least:

$$s_c = max(|V'| - |\bar{N}(V')|)$$

V' = *a set of vertices colored with c*
$\bar{N}(V')$ = *neighborhood with color* $\neq c$

A lower bound for MSV

Proof.

Fix a feasible partition G' and a color c.
Let V' be a set that maximizes s_c.

A lower bound for MSV

Proof.

Fix a feasible partition G' and a color c.
Let V' be a set that maximizes s_c.

For each non-singleton $v' \in V'$ pick a neighbor $n(v')$ in G'.

A lower bound for MSV

Proof.

Fix a feasible partition G' and a color c.
Let V' be a set that maximizes s_c.

For each non-singleton $v' \in V'$ pick a neighbor $n(v')$ in G'.
Any two vertices in V' are in different components of G' $\Rightarrow$ the vertices $n(v')$ are disjoint.

A lower bound for MSV

Proof.

Fix a feasible partition G' and a color c.
Let V' be a set that maximizes s_c.

For each non-singleton $v' \in V'$ pick a neighbor $n(v')$ in G'.
Any two vertices in V' are in different components of G' $\Rightarrow$ the vertices $n(v')$ are disjoint.

$\#$ non-singletons $\leq |\bar{N}(V')|$
$\#$ singletons is at least $|V'| - |\bar{N}(V')|$

□

Exact algorithm for MSV (sketch)

1. Start with an edgeless graph G'

Exact algorithm for MSV (sketch)

1. Start with an edgeless graph G'
2. At each step add an alternating path of odd length

Exact algorithm for MSV (sketch)

1. Start with an edgeless graph G'
2. At each step add an alternating path of odd length
3. The path has edges from G and every second edge is in G'

Exact algorithm for MSV (sketch)

1. Start with an edgeless graph G'
2. At each step add an alternating path of odd length
3. The path has edges from G and every second edge is in G'
4. $\#$ singleton vertices of color c decreases

Exact algorithm for MSV (sketch)

1. Start with an edgeless graph G'
2. At each step add an alternating path of odd length
3. The path has edges from G and every second edge is in G'
4. $\#$ singleton vertices of color c decreases
5. $\#$ other singletons does not increase

Exact algorithm for MSV (sketch)

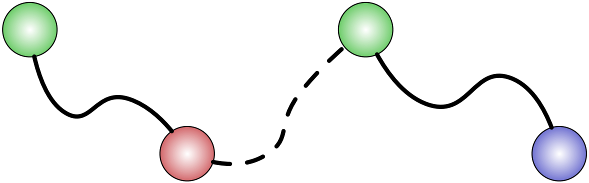
1. Start with an edgeless graph G'
2. At each step add an alternating path of odd length
3. The path has edges from G and every second edge is in G'
4. $\#$ singleton vertices of color c decreases
5. $\#$ other singletons does not increase
6. Each connected component of G' is a singleton, edge or star

Exact algorithm for MSV (sketch)

1. Start with an edgeless graph G'
2. At each step add an alternating path of odd length
3. The path has edges from G and every second edge is in G'
4. $\#$ singleton vertices of color c decreases
5. $\#$ other singletons does not increase
6. Each connected component of G' is a singleton, edge or star
7. When there is no alternating path, we match the lower bound

How to find alternating paths?

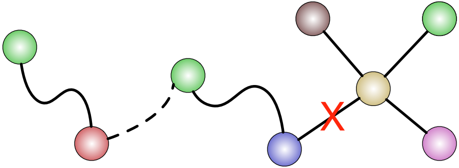
- Start with a singleton of col. c
- Connect a vertex of col. c with a vertex of col. $\neq c$ and viceversa
- Alternate edges in $E \setminus E'$ with edges in E'



Where to end the alternating path?

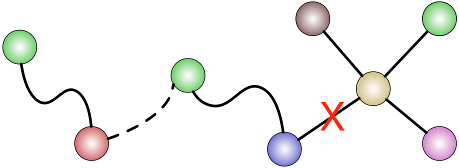
Where to end the alternating path?

- End vertex is a leaf of a star ⇒ remove the leaf from the star

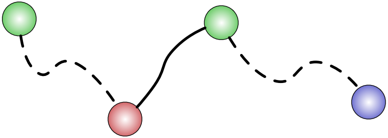


Where to end the alternating path?

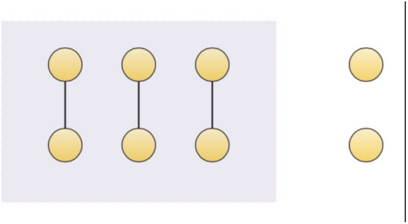
- End vertex is a leaf of a star $\Rightarrow$ remove the leaf from the star



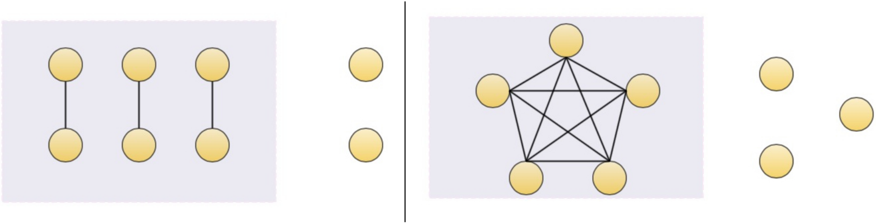
- No col. c in the component $\Rightarrow$ one singleton less + vertices of color c are “switched” between components



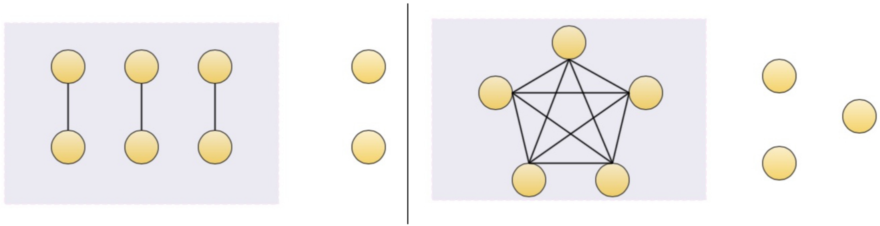
$\sqrt{2 \cdot \mathrm{OPT}}$-approximation



$\sqrt{2 \cdot \mathrm{OPT}}$-approximation



$\sqrt{2 \cdot \mathrm{OPT}}$-approximation

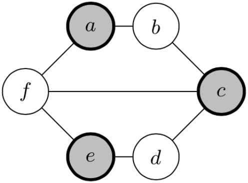


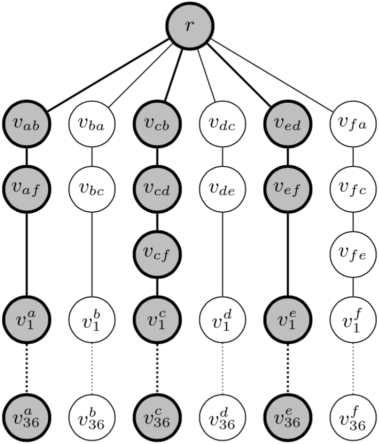
(Handwaving)

In the worst case: $\mathrm{OPT} = \binom{|V|}{2}$ and $\mathrm{MSV} = |V|/2$

Approximation ratio: $\frac{\binom{|V|}{2}}{\frac{|V|}{2}} = \sqrt{2 \cdot \mathrm{OPT}}$

$|V|^{1/3-\varepsilon}$ hardness of approximation for MEC





Hardness of approximating MEC

Lemma

IS of size $\alpha \Rightarrow$ MEC of value $\binom{\alpha n^2}{2}$.

Lemma

MEC of value $n^5/2 + \alpha^2 \cdot n^4 \Rightarrow$ IS of size α.

Theorem

It is NP-hard to approximate MEC within a factor of $|V|^{1/3-\varepsilon}$.

Proof.

NP-hard: IS $< n^\varepsilon$ or IS $\geq n^{1-\varepsilon}$?
$OPT_{MEC} \leq n^5/2 + n^{4+2\varepsilon}$
$OPT_{MEC} \geq n^{6-2\varepsilon}/2$
The number of vertices of G' is in $\Theta(n^3)$ □

APX-hardness of MEC for $|C| = 3$

Reduction from Max Bounded 3D-matching:
each element is in ≤ 3 triples.

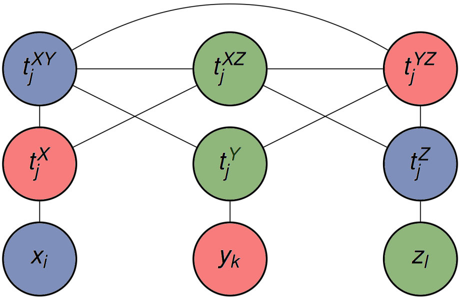


Figure : A subgraph corresponding to a triple $t_j = (x_i, y_k, z_l)$

APX-hardness of MEC for $|C| = 3$

All the vertices are matchend $\Rightarrow OPT_{MEC} = 6\cdot$ # triples + # elements

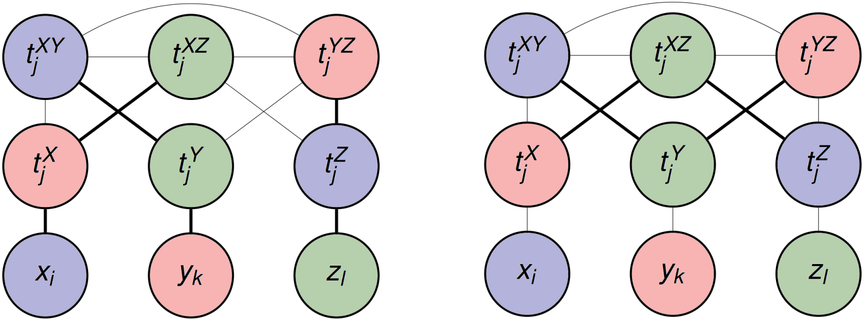


Figure : A triple from T' (left) and a triple from $T \setminus T'$ (right).

Future work

- Other meaningful objective functions.
- Test MSV algorithm on real data.
- Close the gap between the hardness and approximability of MEC.
- Maximizing the number of edges in the connected components.

Merci!

Dziękuję!

Спасибо!

Mulţumesc!

Thank you!