Approximating Shortest Connected Graph Transformation for Trees

SOFSEM 2020

Nicolas Bousquet, Alice Joffard

January 21, 2020







SHORTEST GRAPH TRANSFORMATION (SGT)

SHORTEST GRAPH TRANSFORMATION (SGT)

INPUT: Two multigraphs G,H with the same vertices and the same degree sequence, an integer k

SHORTEST GRAPH TRANSFORMATION (SGT)

INPUT: Two multigraphs G,H with the same vertices and the same degree sequence, an integer k

OUTPUT: True iff there exists a sequence of at most k flips that transforms G into H

SHORTEST GRAPH TRANSFORMATION (SGT)

INPUT: Two multigraphs G,H with the same vertices and the same degree sequence, an integer k

OUTPUT: True iff there exists a sequence of at most k flips that transforms G into H

SHORTEST CONNECTED GRAPH TRANSFORMATION (SCGT)

SHORTEST CONNECTED GRAPH TRANSFORMATION (SCGT)

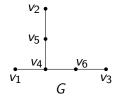
INPUT: Two connected multigraphs G,H with the same vertices and the same degree sequence, an integer k

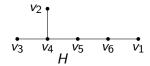
SHORTEST CONNECTED GRAPH TRANSFORMATION (SCGT)

INPUT: Two connected multigraphs G,H with the same vertices and the same degree sequence, an integer k

SHORTEST CONNECTED GRAPH TRANSFORMATION (SCGT)

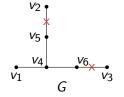
INPUT: Two connected multigraphs G,H with the same vertices and the same degree sequence, an integer k

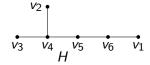




SHORTEST CONNECTED GRAPH TRANSFORMATION (SCGT)

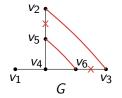
INPUT: Two connected multigraphs G,H with the same vertices and the same degree sequence, an integer k

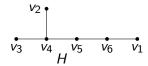




SHORTEST CONNECTED GRAPH TRANSFORMATION (SCGT)

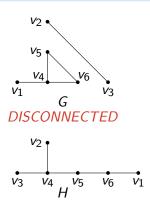
INPUT: Two connected multigraphs G,H with the same vertices and the same degree sequence, an integer k





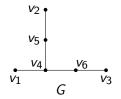
SHORTEST CONNECTED GRAPH TRANSFORMATION (SCGT)

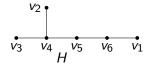
INPUT: Two connected multigraphs G,H with the same vertices and the same degree sequence, an integer k



SHORTEST CONNECTED GRAPH TRANSFORMATION (SCGT)

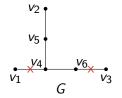
INPUT: Two connected multigraphs G,H with the same vertices and the same degree sequence, an integer k

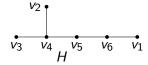




SHORTEST CONNECTED GRAPH TRANSFORMATION (SCGT)

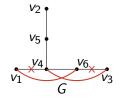
INPUT: Two connected multigraphs G,H with the same vertices and the same degree sequence, an integer k

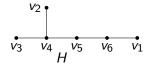




SHORTEST CONNECTED GRAPH TRANSFORMATION (SCGT)

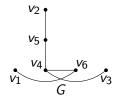
INPUT: Two connected multigraphs G,H with the same vertices and the same degree sequence, an integer k

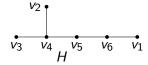




SHORTEST CONNECTED GRAPH TRANSFORMATION (SCGT)

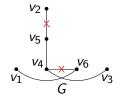
INPUT: Two connected multigraphs G,H with the same vertices and the same degree sequence, an integer k

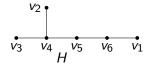




SHORTEST CONNECTED GRAPH TRANSFORMATION (SCGT)

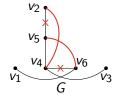
INPUT: Two connected multigraphs G,H with the same vertices and the same degree sequence, an integer k

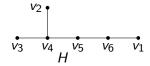




SHORTEST CONNECTED GRAPH TRANSFORMATION (SCGT)

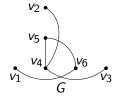
INPUT: Two connected multigraphs G,H with the same vertices and the same degree sequence, an integer k

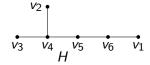




SHORTEST CONNECTED GRAPH TRANSFORMATION (SCGT)

INPUT: Two connected multigraphs G,H with the same vertices and the same degree sequence, an integer k





SORTING BY REVERSALS (SBR)

```
SORTING BY REVERSALS (SBR)
```

INPUT: Two paths P,P' with same vertices and leaves, an integer k

SORTING BY REVERSALS (SBR)

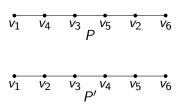
INPUT: Two paths P,P' with same vertices and leaves, an integer k OUTPUT: True iff there exists a sequence of at most k reversals that transforms P into P'

SORTING BY REVERSALS (SBR)

INPUT: Two paths P,P' with same vertices and leaves, an integer k OUTPUT: True iff there exists a sequence of at most k reversals that transforms P into P'

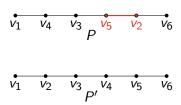
SORTING BY REVERSALS (SBR)

INPUT: Two paths P,P' with same vertices and leaves, an integer k OUTPUT: True iff there exists a sequence of at most k reversals that transforms P into P'



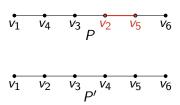
SORTING BY REVERSALS (SBR)

INPUT: Two paths P,P' with same vertices and leaves, an integer k OUTPUT: True iff there exists a sequence of at most k reversals that transforms P into P'



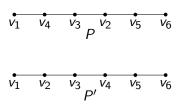
SORTING BY REVERSALS (SBR)

INPUT: Two paths P,P' with same vertices and leaves, an integer k OUTPUT: True iff there exists a sequence of at most k reversals that transforms P into P'



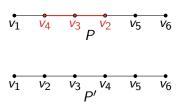
SORTING BY REVERSALS (SBR)

INPUT: Two paths P,P' with same vertices and leaves, an integer k OUTPUT: True iff there exists a sequence of at most k reversals that transforms P into P'



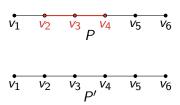
SORTING BY REVERSALS (SBR)

INPUT: Two paths P,P' with same vertices and leaves, an integer k OUTPUT: True iff there exists a sequence of at most k reversals that transforms P into P'



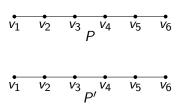
SORTING BY REVERSALS (SBR)

INPUT: Two paths P,P' with same vertices and leaves, an integer k OUTPUT: True iff there exists a sequence of at most k reversals that transforms P into P'



SORTING BY REVERSALS (SBR)

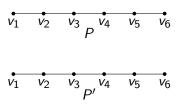
INPUT: Two paths P,P' with same vertices and leaves, an integer k OUTPUT: True iff there exists a sequence of at most k reversals that transforms P into P'



SORTING BY REVERSALS (SBR)

INPUT: Two paths P,P' with same vertices and leaves, an integer k OUTPUT: True iff there exists a sequence of at most k reversals that transforms P into P'

Reversal: Inversion of a subpath

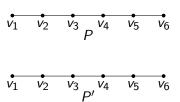


For paths, SCGT is equivalent to SbR

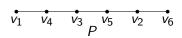
SORTING BY REVERSALS (SBR)

INPUT: Two paths P,P' with same vertices and leaves, an integer k OUTPUT: True iff there exists a sequence of at most k reversals that transforms P into P'

Reversal: Inversion of a subpath



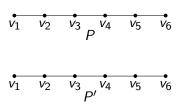
For paths, SCGT is equivalent to SbR



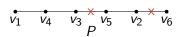
SORTING BY REVERSALS (SBR)

INPUT: Two paths P,P' with same vertices and leaves, an integer k OUTPUT: True iff there exists a sequence of at most k reversals that transforms P into P'

Reversal: Inversion of a subpath



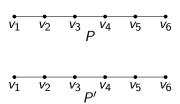
For paths, SCGT is equivalent to SbR

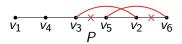


SORTING BY REVERSALS (SBR)

INPUT: Two paths P,P' with same vertices and leaves, an integer k OUTPUT: True iff there exists a sequence of at most k reversals that transforms P into P'

Reversal: Inversion of a subpath

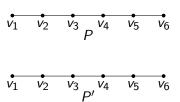


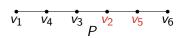


SORTING BY REVERSALS (SBR)

INPUT: Two paths P,P' with same vertices and leaves, an integer k OUTPUT: True iff there exists a sequence of at most k reversals that transforms P into P'

Reversal: Inversion of a subpath

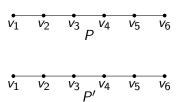


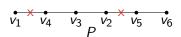


SORTING BY REVERSALS (SBR)

INPUT: Two paths P,P' with same vertices and leaves, an integer k OUTPUT: True iff there exists a sequence of at most k reversals that transforms P into P'

Reversal: Inversion of a subpath

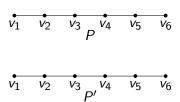


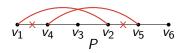


SORTING BY REVERSALS (SBR)

INPUT: Two paths P,P' with same vertices and leaves, an integer k OUTPUT: True iff there exists a sequence of at most k reversals that transforms P into P'

Reversal: Inversion of a subpath

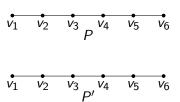


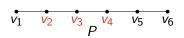


SORTING BY REVERSALS (SBR)

INPUT: Two paths P,P' with same vertices and leaves, an integer k OUTPUT: True iff there exists a sequence of at most k reversals that transforms P into P'

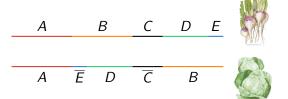
Reversal: Inversion of a subpath



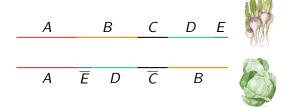


 Application of SbR to the computation of genetic distance between species

 Application of SbR to the computation of genetic distance between species

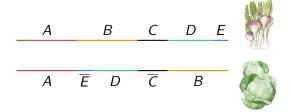


 Application of SbR to the computation of genetic distance between species

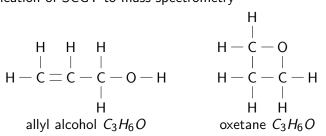


Application of SCGT to mass spectrometry

 Application of SbR to the computation of genetic distance between species



Application of SCGT to mass spectrometry



• SbR:

- SbR:
 - Capraca, 97: SbR is NP-complete

- SbR:
 - Capraca, 97: SbR is NP-complete
 - Berman et al, 02: 1.375-approximation algorithm for SbR

- SbR:
 - Capraca, 97: SbR is NP-complete
 - Berman et al, 02: 1.375-approximation algorithm for SbR
- SGT and SCGT:

- SbR:
 - Capraca, 97: SbR is NP-complete
 - Berman et al, 02: 1.375-approximation algorithm for SbR
- SGT and SCGT:
 - Hakimi, 63: Always a transformation for SGT

- SbR:
 - Capraca, 97: SbR is NP-complete
 - Berman et al, 02: 1.375-approximation algorithm for SbR
- SGT and SCGT:
 - Hakimi, 63: Always a transformation for SGT
 - Will, 99: SGT is NP-complete

- SbR:
 - Capraca, 97: SbR is NP-complete
 - Berman et al, 02: 1.375-approximation algorithm for SbR
- SGT and SCGT:
 - Hakimi, 63: Always a transformation for SGT
 - Will, 99: SGT is NP-complete
 - Bereg and Ito, 16: 1.5-approximation algorithm for SGT

- SbR:
 - Capraca, 97: SbR is NP-complete
 - Berman et al, 02: 1.375-approximation algorithm for SbR
- SGT and SCGT:
 - Hakimi, 63: Always a transformation for SGT
 - Will, 99: SGT is NP-complete
 - Bereg and Ito, 16: 1.5-approximation algorithm for SGT
 - Taylor, 81: Always a transformation for SCGT

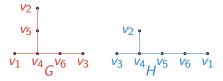
- SbR:
 - Capraca, 97: SbR is NP-complete
 - Berman et al, 02: 1.375-approximation algorithm for SbR
- SGT and SCGT:
 - Hakimi, 63: Always a transformation for SGT
 - Will, 99: SGT is NP-complete
 - Bereg and Ito, 16: 1.5-approximation algorithm for SGT
 - Taylor, 81: Always a transformation for SCGT
 - Bousquet and Mary, 18: 4-approximation algorithm for SCGT

• Distance d(G,H) ($d_c(G,H)$): length of a shortest (connected) transformation between G and H

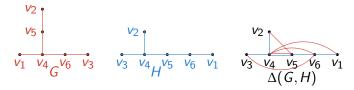
- Distance d(G,H) ($d_c(G,H)$): length of a shortest (connected) transformation between G and H
- Our goal: Approximate $d_c(G, H)$

- Distance d(G,H) ($d_c(G,H)$): length of a shortest (connected) transformation between G and H
- Our goal: Approximate $d_c(G, H)$
- Symmetric difference $\Delta(G,H)$: $(G-H)\cup(H-G)$

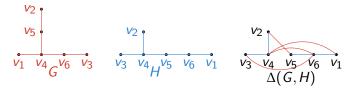
- Distance d(G,H) ($d_c(G,H)$): length of a shortest (connected) transformation between G and H
- Our goal: Approximate $d_c(G, H)$
- Symmetric difference $\Delta(G, H)$: $(G H) \cup (H G)$



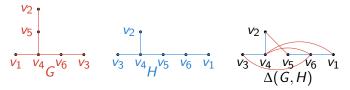
- Distance d(G, H) ($d_c(G, H)$): length of a shortest (connected) transformation between G and H
- Our goal: Approximate $d_c(G, H)$
- Symmetric difference $\Delta(G,H)$: $(G-H) \cup (H-G)$



- Distance d(G,H) ($d_c(G,H)$): length of a shortest (connected) transformation between G and H
- Our goal: Approximate $d_c(G, H)$
- Symmetric difference $\Delta(G,H)$: $(G-H)\cup (H-G)$

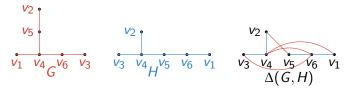


- Distance d(G, H) $(d_c(G, H))$: length of a shortest (connected) transformation between G and H
- Our goal: Approximate $d_c(G, H)$
- Symmetric difference $\Delta(G,H)$: $(G-H)\cup (H-G)$



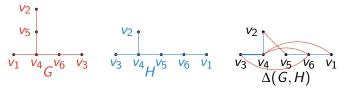
• $d_c(G,H) \geq d(G,H) \geq \frac{|\Delta(G,H)|}{4}$

- Distance d(G,H) ($d_c(G,H)$): length of a shortest (connected) transformation between G and H
- Our goal: Approximate $d_c(G, H)$
- Symmetric difference $\Delta(G,H)$: $(G-H) \cup (H-G)$



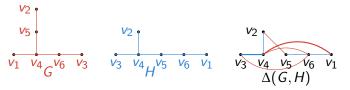
- $d_c(G,H) \geq d(G,H) \geq \frac{|\Delta(G,H)|}{4}$
- Will, 99: $d(G,H) = \frac{|\Delta(G,H)|}{2} mnc(G,H)$ where mnc(G,H) is the maximum number of cycles in a partition of $\Delta(G,H)$ into alternating cycles.

- Distance d(G,H) ($d_c(G,H)$): length of a shortest (connected) transformation between G and H
- Our goal: Approximate $d_c(G, H)$
- Symmetric difference $\Delta(G,H)$: $(G-H) \cup (H-G)$



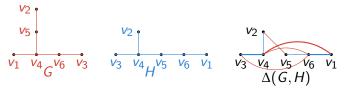
- $d_c(G,H) \geq d(G,H) \geq \frac{|\Delta(G,H)|}{4}$
- Will, 99: $d(G,H) = \frac{|\Delta(G,H)|}{2} mnc(G,H)$ where mnc(G,H) is the maximum number of cycles in a partition of $\Delta(G,H)$ into alternating cycles.

- Distance d(G, H) ($d_c(G, H)$): length of a shortest (connected) transformation between G and H
- Our goal: Approximate $d_c(G, H)$
- Symmetric difference $\Delta(G,H)$: $(G-H) \cup (H-G)$



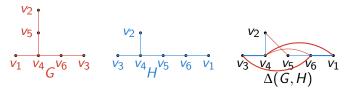
- $d_c(G,H) \geq d(G,H) \geq \frac{|\Delta(G,H)|}{4}$
- Will, 99: $d(G,H) = \frac{|\Delta(G,H)|}{2} mnc(G,H)$ where mnc(G,H) is the maximum number of cycles in a partition of $\Delta(G,H)$ into alternating cycles.

- Distance d(G, H) ($d_c(G, H)$): length of a shortest (connected) transformation between G and H
- Our goal: Approximate $d_c(G, H)$
- Symmetric difference $\Delta(G,H)$: $(G-H) \cup (H-G)$



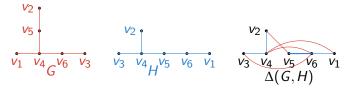
- $d_c(G,H) \geq d(G,H) \geq \frac{|\Delta(G,H)|}{4}$
- Will, 99: $d(G,H) = \frac{|\Delta(G,H)|}{2} mnc(G,H)$ where mnc(G,H) is the maximum number of cycles in a partition of $\Delta(G,H)$ into alternating cycles.

- Distance d(G, H) ($d_c(G, H)$): length of a shortest (connected) transformation between G and H
- Our goal: Approximate $d_c(G, H)$
- Symmetric difference $\Delta(G,H)$: $(G-H) \cup (H-G)$



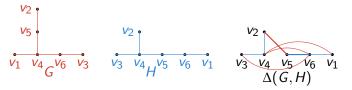
- $d_c(G,H) \geq d(G,H) \geq \frac{|\Delta(G,H)|}{4}$
- Will, 99: $d(G,H) = \frac{|\Delta(G,H)|}{2} mnc(G,H)$ where mnc(G,H) is the maximum number of cycles in a partition of $\Delta(G,H)$ into alternating cycles.

- Distance d(G, H) ($d_c(G, H)$): length of a shortest (connected) transformation between G and H
- Our goal: Approximate $d_c(G, H)$
- Symmetric difference $\Delta(G,H)$: $(G-H) \cup (H-G)$



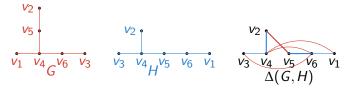
- $d_c(G,H) \geq d(G,H) \geq \frac{|\Delta(G,H)|}{4}$
- Will, 99: $d(G,H) = \frac{|\Delta(G,H)|}{2} mnc(G,H)$ where mnc(G,H) is the maximum number of cycles in a partition of $\Delta(G,H)$ into alternating cycles.

- Distance d(G, H) ($d_c(G, H)$): length of a shortest (connected) transformation between G and H
- Our goal: Approximate $d_c(G, H)$
- Symmetric difference $\Delta(G,H)$: $(G-H) \cup (H-G)$



- $d_c(G,H) \geq d(G,H) \geq \frac{|\Delta(G,H)|}{4}$
- Will, 99: $d(G,H) = \frac{|\Delta(G,H)|}{2} mnc(G,H)$ where mnc(G,H) is the maximum number of cycles in a partition of $\Delta(G,H)$ into alternating cycles.

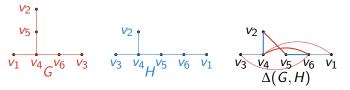
- Distance d(G,H) ($d_c(G,H)$): length of a shortest (connected) transformation between G and H
- Our goal: Approximate $d_c(G, H)$
- Symmetric difference $\Delta(G,H)$: $(G-H) \cup (H-G)$



- $d_c(G,H) \geq d(G,H) \geq \frac{|\Delta(G,H)|}{4}$
- Will, 99: $d(G,H) = \frac{|\Delta(G,H)|}{2} mnc(G,H)$ where mnc(G,H) is the maximum number of cycles in a partition of $\Delta(G,H)$ into alternating cycles.

Preliminaries

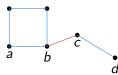
- Distance d(G,H) ($d_c(G,H)$): length of a shortest (connected) transformation between G and H
- Our goal: Approximate $d_c(G, H)$
- Symmetric difference $\Delta(G,H)$: $(G-H) \cup (H-G)$

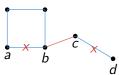


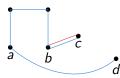
- $d_c(G,H) \geq d(G,H) \geq \frac{|\Delta(G,H)|}{4}$
- Will, 99: $d(G,H) = \frac{|\Delta(G,H)|}{2} mnc(G,H)$ where mnc(G,H) is the maximum number of cycles in a partition of $\Delta(G,H)$ into alternating cycles.

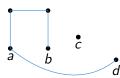




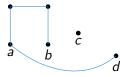




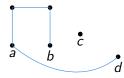


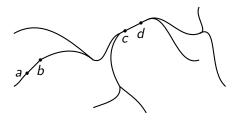


• If an edge ab of G-H is in a cycle of G, we reduce $|\Delta(G,H)|$ by 2 in 1 flip

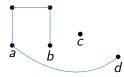


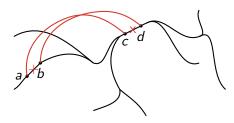
• If an edge ab of G-H is in a cycle of G, we reduce $|\Delta(G,H)|$ by 2 in 1 flip



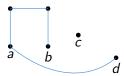


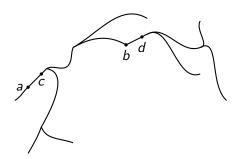
• If an edge ab of G-H is in a cycle of G, we reduce $|\Delta(G,H)|$ by 2 in 1 flip





• If an edge ab of G-H is in a cycle of G, we reduce $|\Delta(G,H)|$ by 2 in 1 flip





Theorem: SCGT admits a 2.5-approximation algorithm.

Theorem: SCGT admits a 2.5-approximation algorithm.

Theorem: SCGT admits a 2.5-approximation algorithm.

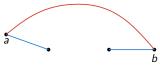
• Upper bound: Reduce $|\Delta(G, H)|$ by at least 4 in at most 3 flips

ă

Theorem: SCGT admits a 2.5-approximation algorithm.



Theorem: SCGT admits a 2.5-approximation algorithm.



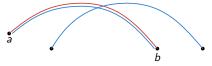
Theorem: SCGT admits a 2.5-approximation algorithm.



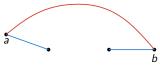
Theorem: SCGT admits a 2.5-approximation algorithm.



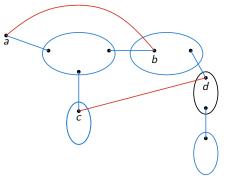
Theorem: SCGT admits a 2.5-approximation algorithm.



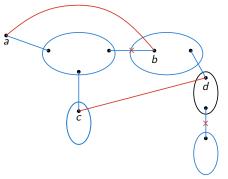
Theorem: SCGT admits a 2.5-approximation algorithm.



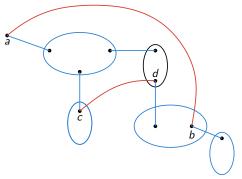
Theorem: SCGT admits a 2.5-approximation algorithm.



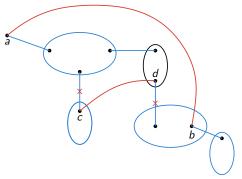
Theorem: SCGT admits a 2.5-approximation algorithm.



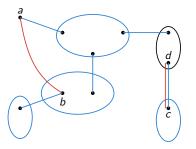
Theorem: SCGT admits a 2.5-approximation algorithm.



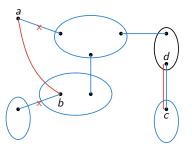
Theorem: SCGT admits a 2.5-approximation algorithm.



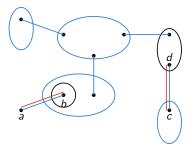
Theorem: SCGT admits a 2.5-approximation algorithm.



Theorem: SCGT admits a 2.5-approximation algorithm.

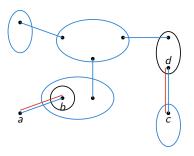


Theorem: SCGT admits a 2.5-approximation algorithm.



Theorem: SCGT admits a 2.5-approximation algorithm.

• Upper bound: Reduce $|\Delta(G, H)|$ by at least 4 in at most 3 flips



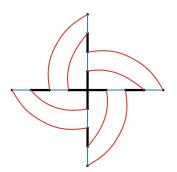
• Lower bound: $d_c(G,H) \ge \frac{|\Delta(G,H)|}{2} - mnc(G,H)$

• There exist G_k and H_k for which if we only flip edges of the symmetric difference,

$$d_c(G_k, H_k) \ge \frac{3}{2} \left(\frac{|\Delta(G_k, H_k)|}{2} - mnc(G_k, H_k) \right)$$

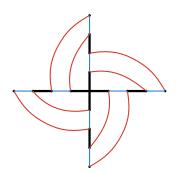
• There exist G_k and H_k for which if we only flip edges of the symmetric difference,

$$d_c(G_k, H_k) \ge \frac{3}{2} \left(\frac{|\Delta(G_k, H_k)|}{2} - mnc(G_k, H_k) \right)$$



 There exist G_k and H_k for which if we only flip edges of the symmetric difference,

$$d_c(G_k, H_k) \ge \frac{3}{2} \left(\frac{|\Delta(G_k, H_k)|}{2} - mnc(G_k, H_k) \right)$$



 With this lower bound, can not do better than a 1.5-approximation

• Conjecture: The shortest transformation from G_k to H_k has length $\frac{|\Delta(G_k,H_k)|}{2}-1$

- Conjecture: The shortest transformation from G_k to H_k has length $\frac{|\Delta(G_k,H_k)|}{2}-1$
 - \rightarrow True when we only flip edges of the same cycle of the decomposition of $\Delta(G,H)$ into alternating cycles

- Conjecture: The shortest transformation from G_k to H_k has length $\frac{|\Delta(G_k,H_k)|}{2}-1$
 - \rightarrow True when we only flip edges of the same cycle of the decomposition of $\Delta(G,H)$ into alternating cycles
- A shortest transformation from G to H only flips edges of $\Delta(G,H)$?

- Conjecture: The shortest transformation from G_k to H_k has length $\frac{|\Delta(G_k,H_k)|}{2}-1$
 - \rightarrow True when we only flip edges of the same cycle of the decomposition of $\Delta(G,H)$ into alternating cycles
- A shortest transformation from G to H only flips edges of $\Delta(G,H)$?
 - \rightarrow Open for paths

- Conjecture: The shortest transformation from G_k to H_k has length $\frac{|\Delta(G_k,H_k)|}{2}-1$
 - \rightarrow True when we only flip edges of the same cycle of the decomposition of $\Delta(G,H)$ into alternating cycles
- A shortest transformation from G to H only flips edges of $\Delta(G,H)$?
 - ightarrow Open for paths
- Find a better lower bound to improve our approximation ratio

- Conjecture: The shortest transformation from G_k to H_k has length $\frac{|\Delta(G_k,H_k)|}{2}-1$
 - \rightarrow True when we only flip edges of the same cycle of the decomposition of $\Delta(G,H)$ into alternating cycles
- A shortest transformation from G to H only flips edges of Δ(G, H)?
 - ightarrow Open for paths
- Find a better lower bound to improve our approximation ratio

