The r-Gather Problem in Euclidean Space

Esther M. Arkin[a], Jie Gao[a], Matthew P. Johnson[b],
Joseph S. B. Mitchell[a], and Jiemin Zeng[a]

a Stony Brook University b Lehman College and the Graduate Center, CUNY

October 24, 2015

Outline

1. Intro
2. Hardness
   - $r \geq 4$ and cost $=$ max bounding disk diameter
3. Algorithms
4. Conclusion

Outline

1. Intro
2. Hardness
   - $r \geq 4$ and cost $=$ max bounding disk diameter
3. Algorithms
4. Conclusion

Scenario: facility location-like "gathering" (clustering) problem

Given: collection P of points in a metric space

Scenario: facility location-like "gathering" (clustering) problem

Given: collection P of points in a metric space

Solution: partition of P into clusters $C_1, .., C_k$

Scenario: facility location-like “gathering” (clustering) problem

Given: collection P of points in a metric space

Solution: partition of P into clusters $C_1, .., C_k$

Requirement: $|C_i| \geq r \ \ \forall i$

Scenario: facility location-like "gathering" (clustering) problem

Given: collection P of points in a metric space

Solution: partition of P into clusters $C_1, .., C_k$

Requirement: $|C_i| \geq r \ \ \forall i$

for each cluster C_i, let d_i be the diameter of its bounding disk

Scenario: facility location-like "gathering" (clustering) problem

Given: collection P of points in a metric space

Solution: partition of P into clusters $C_1, .., C_k$

Requirement: $|C_i| \geq r \;\; \forall i$

for each cluster C_i*, let* d_i *be the diameter of its bounding disk*

Solution cost: $\max_i d_i$

Scenario: facility location-like "gathering" (clustering) problem

Intuitively, each cluster's facility is only worth opening if it's going to do $\geq r$ points.

Scenario: facility location-like "gathering" (clustering) problem

Intuitively, each cluster's facility is only worth opening if it's going to do $\geq r$ points.

You want to minimize the (max) cost of the disks that cover them.

Scenario: facility location-like "gathering" (clustering) problem

Intuitively, each cluster's facility is only worth opening if it's going to do $\geq r$ points.

You want to minimize the (max) cost of the disks that cover them.

Points included in multiple disks can be assigned arbitrarily.

Scenario: facility location-like "gathering" (clustering) problem

Intuitively, each cluster's facility is only worth opening if it's going to do $\geq r$ points.

You want to minimize the (max) cost of the disks that cover them.

Points included in multiple disks can be assigned arbitrarily.

The r is **requirement**, not radius.

Prior work

- Optimally solvable for $r = 2$.

Prior work

- Optimally solvable for $r = 2$.
- 2-approximable for $r > 2$.

Prior work

- Optimally solvable for $r = 2$.
- 2-approximable for $r > 2$.
- Known to be hard to approx better than factor 2 for $r > 2$.

Privacy motivation r-anonymity

One motivation: privacy of mobile device locations

Privacy motivation r-anonymity

One motivation: privacy of mobile device locations

One approach: coarsen location values

Privacy motivation r-anonymity

One motivation: privacy of mobile device locations

One approach: coarsen location values

Another: group points into cluster so each point appears with $\geq k$ others, i.e. k-anonymity
(here $r = k$)

Privacy motivation r-anonymity

One motivation: privacy of mobile device locations

One approach: coarsen location values

Another: group points into cluster so each point appears with $\geq k$ others, i.e. k-anonymity
(here $r = k$)

In this application, the space is arguably (?) Euclidean

Euclidean special case

Our question: can you do better in the Euclidean plane?

Euclidean special case

Our question: can you do better in the Euclidean plane?

We give partial (negative) answers: hardness of approx lower bounds, whose values depend on r (but are < 2).

Euclidean special case

Our question: can you do better in the Euclidean plane?

We give partial (negative) answers: hardness of approx lower bounds, whose values depend on r (but are < 2).

Also, stronger lower bounds for an alt df of diameter.

Euclidean special case

Our question: can you do better in the Euclidean plane?

We give partial (negative) answers: hardness of approx lower bounds, whose values depend on r (but are < 2).

Also, stronger lower bounds for an alt df of diameter.

Reductions from Planar 3SAT and Planar Circuit 3SAT.

Results

We obtain these hardness of approx factors:

Results

We obtain these hardness of approx factors:

	$r = 2$	$r = 3$	$r = 4$	$r \geq 5$
disk diam	(in P)	$\sqrt{13}/2 \approx 1.802$	$\frac{\sqrt{35}+\sqrt{3}}{4} \approx 1.912$	same
pair dist	(in P)	$\sqrt{2+\sqrt{3}} \approx 1.931$	same	2

Outline

1. Intro
2. Hardness
   - $r \geq 4$ and cost $=$ max bounding disk diameter
3. Algorithms
4. Conclusion

Hardness

Construction:

Hardness

Construction:

- Vars and clauses connected by chains of unit-diam disks.

Hardness

Construction:

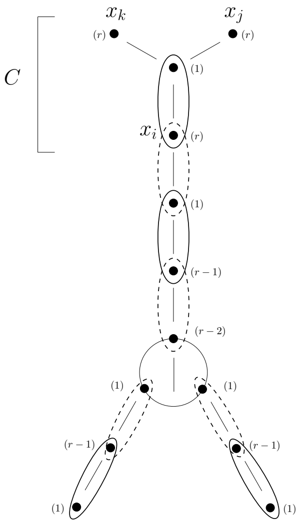
- Vars and clauses connected by chains of unit-diam disks.
- Vars = cyclic chain of unit-diam disks.

Hardness

Construction:

- Vars and clauses connected by chains of unit-diam disks.
- Vars = cyclic chain of unit-diam disks.
- Clauses: meeting point of three chains — one of them must cover it.

Hardness



Hardness

We show that:

- pos SAT $\rightarrow$ instance with OPT=1
- neg SAT $\rightarrow$ instance with OPT $\geq c$ (for some approx factor $c \in [1, 2]$)

Hardness

We show that:

- pos SAT $\rightarrow$ instance with OPT=1
- neg SAT $\rightarrow$ instance with OPT $\geq c$ (for some approx factor $c \in [1, 2]$)

If SAT is neg, how will the instance "correct itself"?

- correcting within chain would cost 2
- correcting at a clause would cost 2 (or would it!)

Hardness

We show that:

- pos SAT $\to$ instance with OPT=1
- neg SAT $\to$ instance with OPT $\geq c$ (for some approx factor $c \in [1, 2]$)

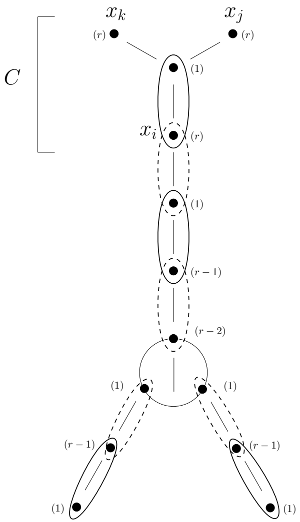
If SAT is neg, how will the instance "correct itself"?

- correcting within chain would cost 2
- correcting at a clause would cost 2 (or would it!)

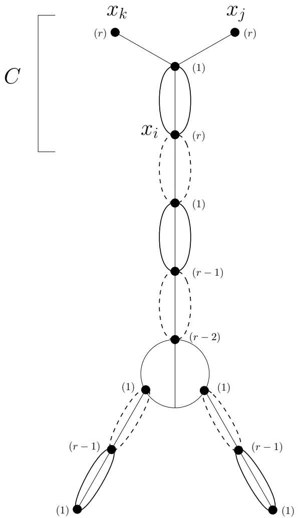
Only other option:

- correcting at a splitter

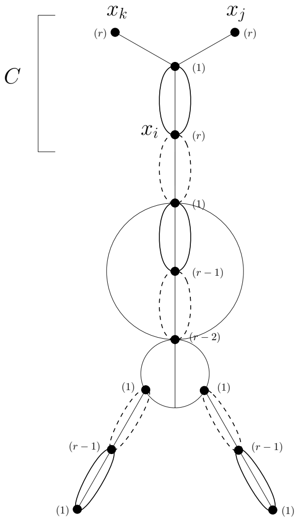
Hardness



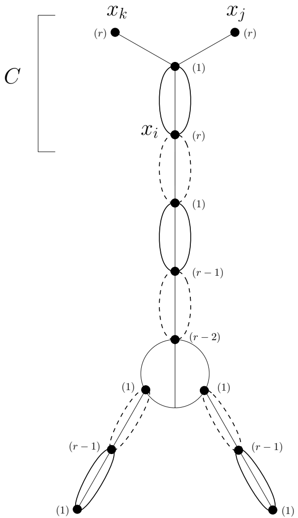
Hardness



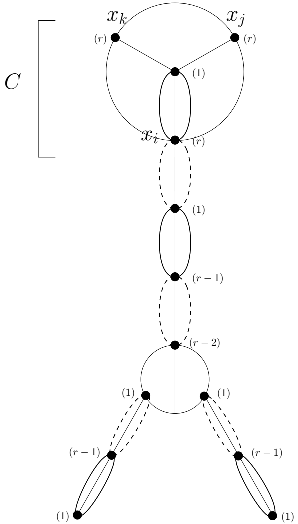
Hardness



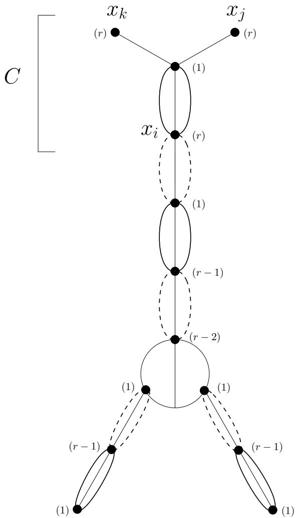
Hardness



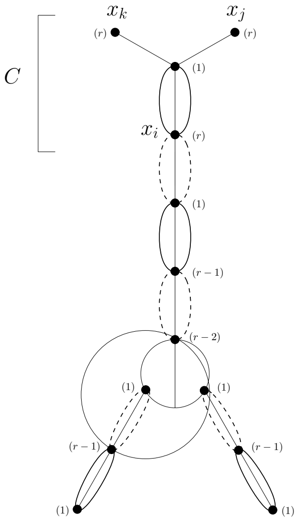
Hardness



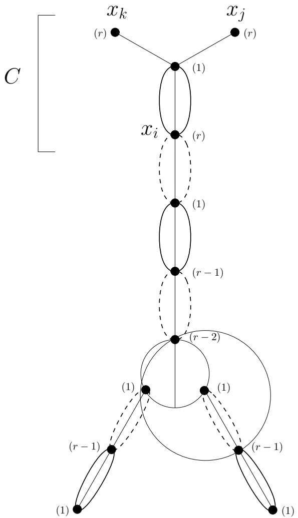
Hardness



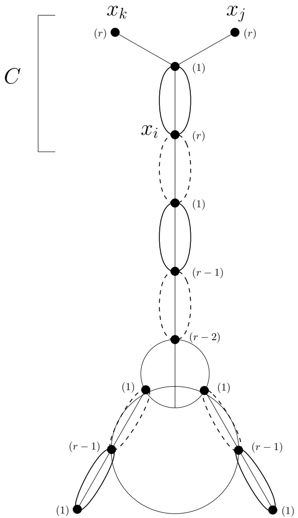
Hardness



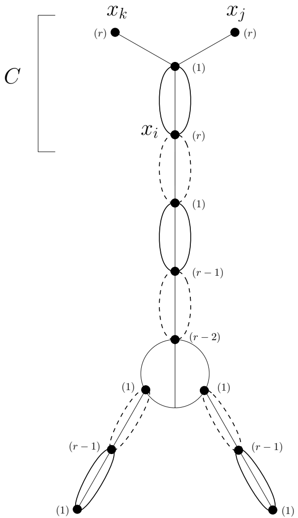
Hardness



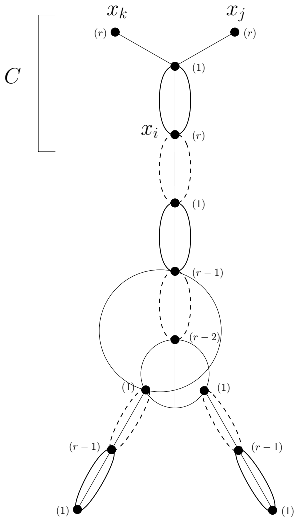
Hardness



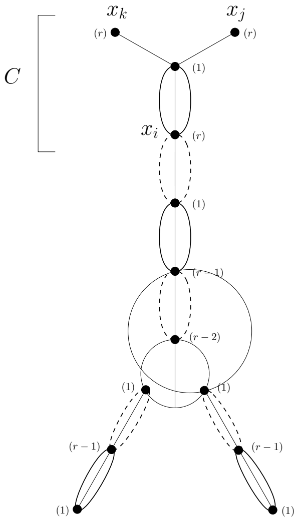
Hardness



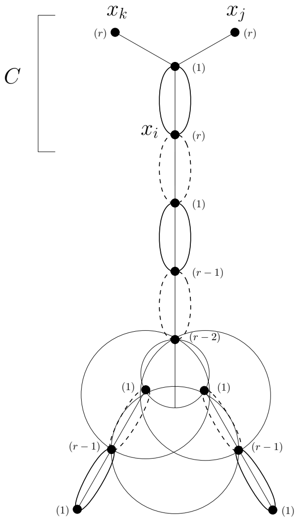
Hardness



Hardness



Hardness



Hardness

How to correct at the splitter?

Hardness

How to correct at the splitter?

wlog, restrict ourselves to a certain configuration with two disks touching a third disk.

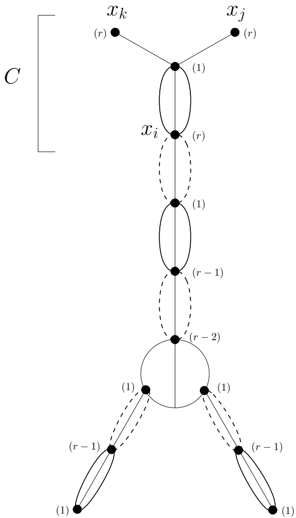
Hardness

How to correct at the splitter?

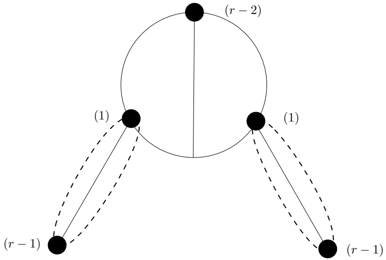
wlog, restrict ourselves to a certain configuration with two disks touching a third disk.

Then we compute optimal way for them to touch it $\rightarrow$ the 1.912 factor.

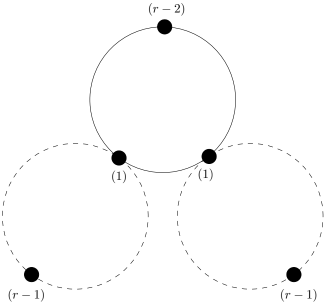
Hardness



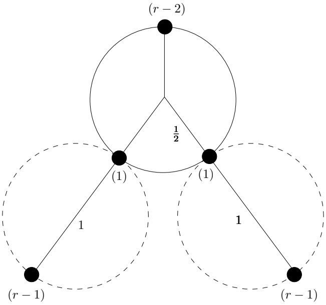
Hardness



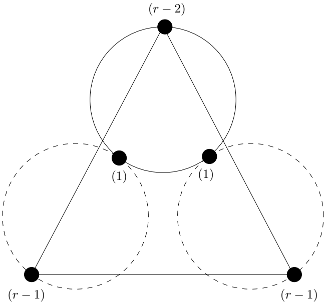
Hardness



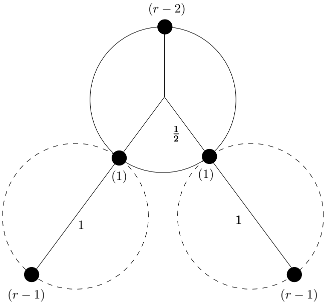
Hardness



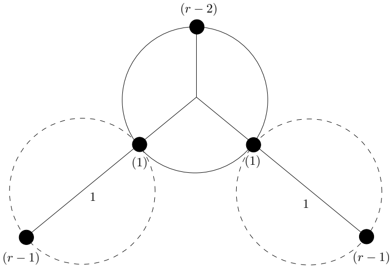
Hardness



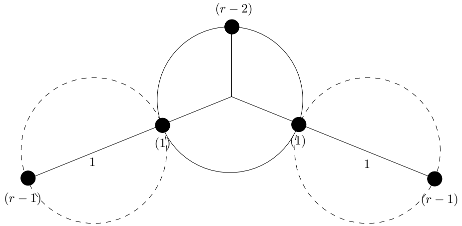
Hardness



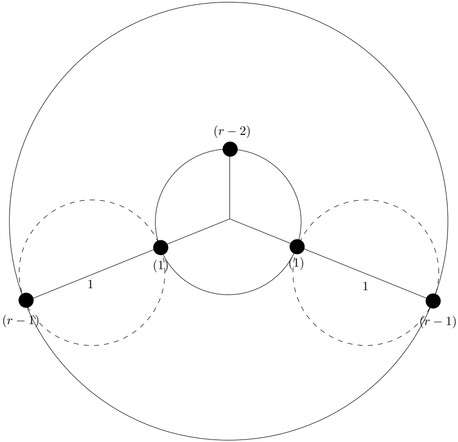
Hardness



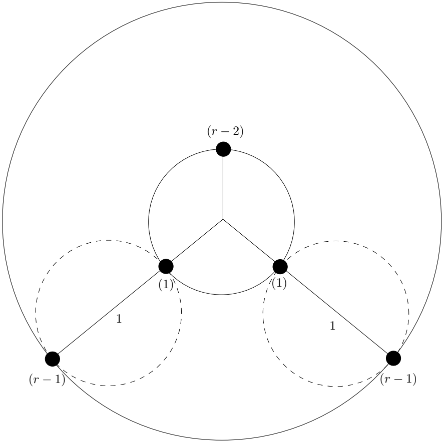
Hardness



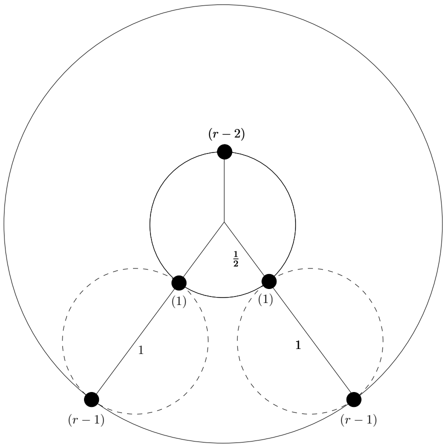
Hardness



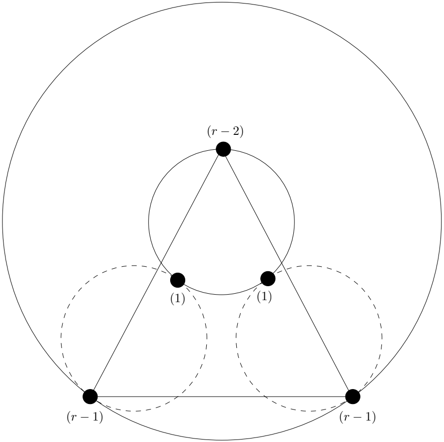
Hardness



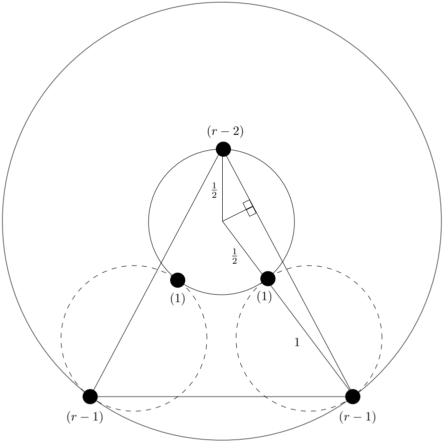
Hardness



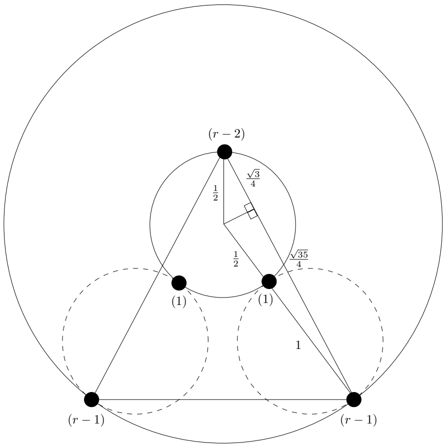
Hardness



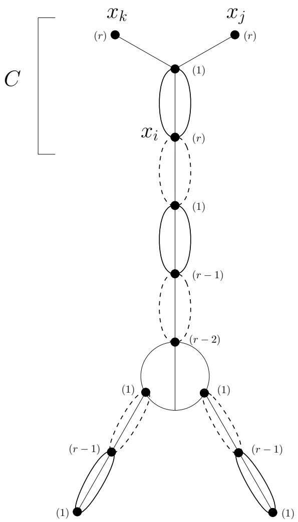
Hardness



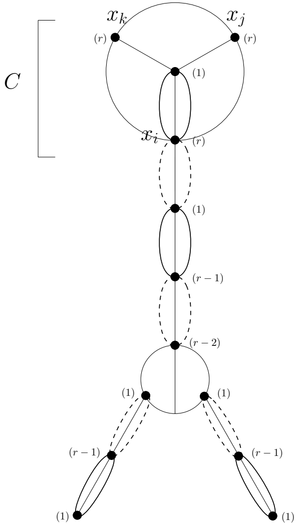
Hardness



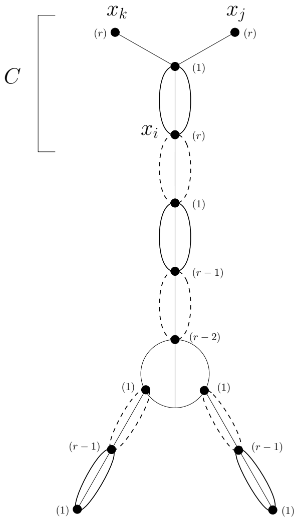
What about $r = 3$?



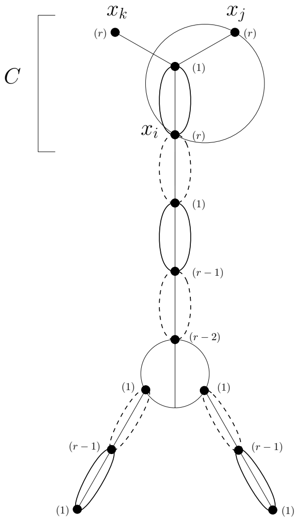
What about $r = 3$?



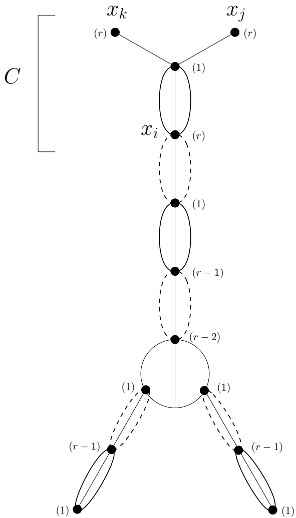
What about $r = 3$?



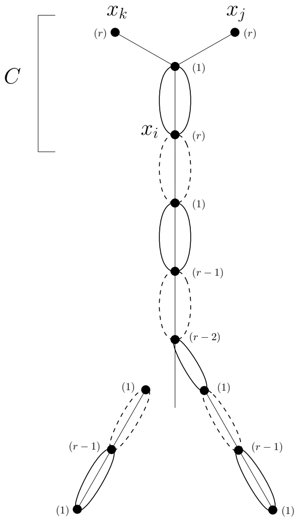
What about $r = 3$?



What about $r = 2$?



What about $r = 2$?



Outline

Algorithms

Algorithms

Nope!

Outline

1. Intro
2. Hardness
   - $r \geq 4$ and cost $=$ max bounding disk diameter
3. Algorithms
4. Conclusion

Conclusion

So: restricting from metric to Euclidean plane, achieved hardness of approx approaching 2:

Conclusion

So: restricting from metric to Euclidean plane, achieved hardness of approx approaching 2:

$\sqrt{3} \approx 1.73$

Conclusion

So: restricting from metric to Euclidean plane, achieved hardness of approx approaching 2:

$\sqrt{3} \approx 1.73 \quad \rightarrow$

Conclusion

So: restricting from metric to Euclidean plane, achieved hardness of approx approaching 2:

$$\sqrt{3} \approx 1.73 \quad \rightarrow \quad 1 + \frac{\sqrt{3}}{2} \approx 1.866$$

Conclusion

So: restricting from metric to Euclidean plane, achieved hardness of approx approaching 2:

$$\sqrt{3} \approx 1.73 \quad \rightarrow \quad 1 + \frac{\sqrt{3}}{2} \approx 1.866 \quad \rightarrow$$

Conclusion

So: restricting from metric to Euclidean plane, achieved hardness of approx approaching 2:

$$\sqrt{3} \approx 1.73 \quad \rightarrow \quad 1 + \frac{\sqrt{3}}{2} \approx 1.866 \quad \rightarrow \quad \frac{\sqrt{3}+\sqrt{35}}{4} \approx 1.912$$

Conclusion

So: restricting from metric to Euclidean plane, achieved hardness of approx approaching 2:

$$\sqrt{3} \approx 1.73 \quad \rightarrow \quad 1 + \frac{\sqrt{3}}{2} \approx 1.866 \quad \rightarrow \quad \frac{\sqrt{3}+\sqrt{35}}{4} \approx 1.912 \quad \not\rightarrow \quad 2$$

Conclusion

So: restricting from metric to Euclidean plane, achieved hardness of approx approaching 2:

$$\sqrt{3} \approx 1.73 \quad \rightarrow \quad 1 + \frac{\sqrt{3}}{2} \approx 1.866 \quad \rightarrow \quad \frac{\sqrt{3}+\sqrt{35}}{4} \approx 1.912 \quad \not\rightarrow \quad 2$$

But gap remains, and no improvement in positive results

Conclusion

So: restricting from metric to Euclidean plane, achieved hardness of approx approaching 2:

$$\sqrt{3} \approx 1.73 \quad \rightarrow \quad 1 + \frac{\sqrt{3}}{2} \approx 1.866 \quad \rightarrow \quad \frac{\sqrt{3}+\sqrt{35}}{4} \approx 1.912 \quad \not\rightarrow \quad 2$$

But gap remains, and no improvement in positive results

Complimentary optimization goal: maximize r, given hard constraint on diameter d

Conclusion

So: restricting from metric to Euclidean plane, achieved hardness of approx approaching 2:

$$\sqrt{3} \approx 1.73 \quad \rightarrow \quad 1 + \frac{\sqrt{3}}{2} \approx 1.866 \quad \rightarrow \quad \frac{\sqrt{3}+\sqrt{35}}{4} \approx 1.912 \quad \not\rightarrow \quad 2$$

But gap remains, and no improvement in positive results

Complimentary optimization goal: maximize r, given hard constraint on diameter d

- hardness proof for $r = 3$ case $\rightarrow$ hard to approx for $c_r < \frac{2}{3}$
- *no known approx alg*

Conclusion

So: restricting from metric to Euclidean plane, achieved hardness of approx approaching 2:

$$\sqrt{3} \approx 1.73 \quad \rightarrow \quad 1 + \frac{\sqrt{3}}{2} \approx 1.866 \quad \rightarrow \quad \frac{\sqrt{3}+\sqrt{35}}{4} \approx 1.912 \quad \not\rightarrow \quad 2$$

But gap remains, and no improvement in positive results

Complimentary optimization goal: maximize r, given hard constraint on diameter d

- hardness proof for $r = 3$ case $\rightarrow$ hard to approx for $c_r < \frac{2}{3}$
- *no known approx alg*

$\rightarrow$ bicriteria problem for (d, r)

Conclusion

So: restricting from metric to Euclidean plane, achieved hardness of approx approaching 2:

$$\sqrt{3} \approx 1.73 \quad \rightarrow \quad 1 + \frac{\sqrt{3}}{2} \approx 1.866 \quad \rightarrow \quad \frac{\sqrt{3}+\sqrt{35}}{4} \approx 1.912 \quad \not\rightarrow \quad 2$$

But gap remains, and no improvement in positive results

Complimentary optimization goal: maximize r, given hard constraint on diameter d

- hardness proof for $r = 3$ case $\rightarrow$ hard to approx for $c_r < \frac{2}{3}$
- *no known approx alg*

$\rightarrow$ bicriteria problem for (d, r)

- Original 2-approx here: $(2, 1)$

Conclusion

So: restricting from metric to Euclidean plane, achieved hardness of approx approaching 2:

$$\sqrt{3} \approx 1.73 \quad \rightarrow \quad 1 + \frac{\sqrt{3}}{2} \approx 1.866 \quad \rightarrow \quad \frac{\sqrt{3}+\sqrt{35}}{4} \approx 1.912 \quad \not\rightarrow \quad 2$$

But gap remains, and no improvement in positive results

Complimentary optimization goal: maximize r, given hard constraint on diameter d

- hardness proof for $r = 3$ case $\rightarrow$ hard to approx for $c_r < \frac{2}{3}$
- *no known approx alg*

$\rightarrow$ bicriteria problem for (d, r)

- Original 2-approx here: $(2, 1)$
- Alt goal: $(1, c_r)$ for some $c_r \in (0, 1)$

Conclusion

So: restricting from metric to Euclidean plane, achieved hardness of approx approaching 2:

$$\sqrt{3} \approx 1.73 \quad \rightarrow \quad 1 + \frac{\sqrt{3}}{2} \approx 1.866 \quad \rightarrow \quad \frac{\sqrt{3}+\sqrt{35}}{4} \approx 1.912 \quad \not\rightarrow \quad 2$$

But gap remains, and no improvement in positive results

Complimentary optimization goal: maximize r, given hard constraint on diameter d

- hardness proof for $r = 3$ case $\rightarrow$ hard to approx for $c_r < \frac{2}{3}$
- *no known approx alg*

$\rightarrow$ bicriteria problem for (d, r)

- Original 2-approx here: $(2, 1)$
- Alt goal: $(1, c_r)$ for some $c_r \in (0, 1)$ (as yet unsuccessful)

Conclusion

So: restricting from metric to Euclidean plane, achieved hardness of approx approaching 2:

$$\sqrt{3} \approx 1.73 \quad \to \quad 1 + \frac{\sqrt{3}}{2} \approx 1.866 \quad \to \quad \frac{\sqrt{3}+\sqrt{35}}{4} \approx 1.912 \quad \not\to \quad 2$$

But gap remains, and no improvement in positive results

Complimentary optimization goal: maximize r, given hard constraint on diameter d

- hardness proof for $r = 3$ case $\to$ hard to approx for $c_r < \frac{2}{3}$
- *no known approx alg*

$\to$ bicriteria problem for (d, r)

- Original 2-approx here: $(2, 1)$
- Alt goal: $(1, c_r)$ for some $c_r \in (0, 1)$ (as yet unsuccessful)

- More modest goal: (c_d, c_r) for some $c_r \in (0, 1)$ and some $c_d < 2$

Conclusion

So: restricting from metric to Euclidean plane, achieved hardness of approx approaching 2:

$$\sqrt{3} \approx 1.73 \quad \rightarrow \quad 1 + \frac{\sqrt{3}}{2} \approx 1.866 \quad \rightarrow \quad \frac{\sqrt{3}+\sqrt{35}}{4} \approx 1.912 \quad \not\rightarrow \quad 2$$

But gap remains, and no improvement in positive results

Complimentary optimization goal: maximize r, given hard constraint on diameter d

- hardness proof for $r = 3$ case $\rightarrow$ hard to approx for $c_r < \frac{2}{3}$
- *no known approx alg*

$\rightarrow$ bicriteria problem for (d, r)

- Original 2-approx here: $(2, 1)$
- Alt goal: $(1, c_r)$ for some $c_r \in (0, 1)$ (as yet unsuccessful)

- More modest goal: (c_d, c_r) for some $c_r \in (0, 1)$ and some $c_d < 2$ (also unsuccessful!!)

Conclusion

	$r=2$	$r=3$	$r=4$	$r\geq 5$
disk diam	(in P)	$\sqrt{13}/2 \approx 1.802$	$\frac{\sqrt{35}+\sqrt{3}}{4} \approx 1.912$	same
pair dist	(in P)	$\sqrt{2+\sqrt{3}} \approx 1.931$	same	2

Thanks!
mpjohnson@gmail.com