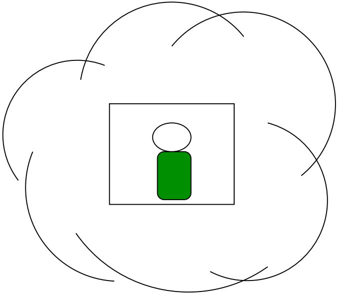
Capacitated Metric Labeling

Ankur Moitra, MIT

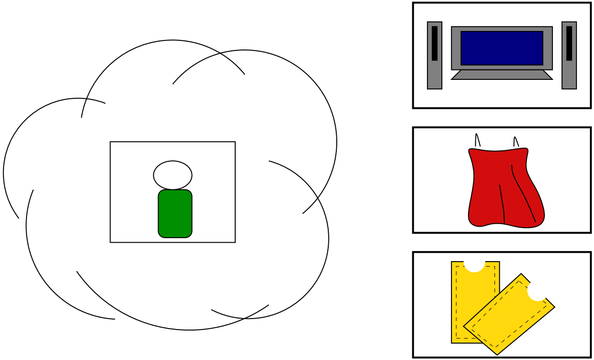
joint work with Matthew Andrews, MohammadTaghi Hajiaghayi and Howard Karloff

January 24, 2011

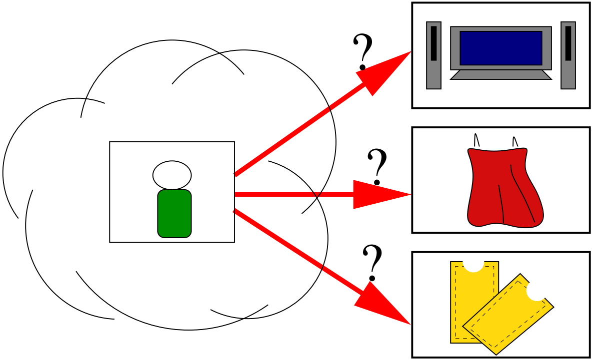
Classification (on a Social Network)



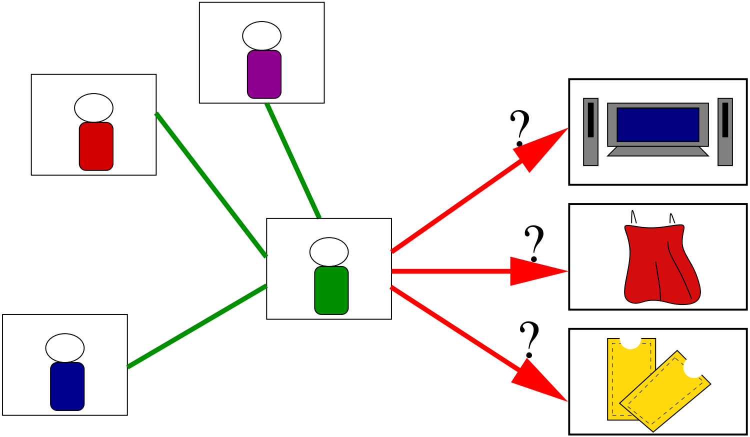
Classification (on a Social Network)



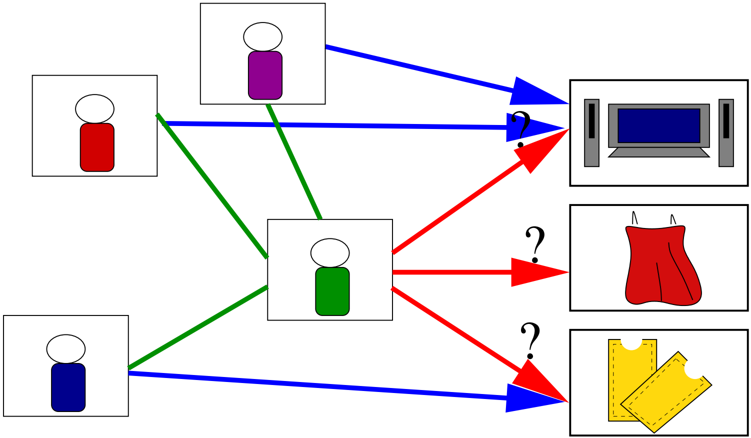
Classification (on a Social Network)



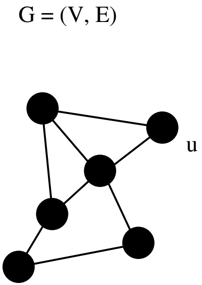
Classification (on a Social Network)



Classification (on a Social Network)

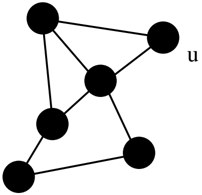


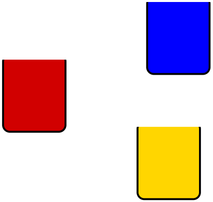
Metric Labeling



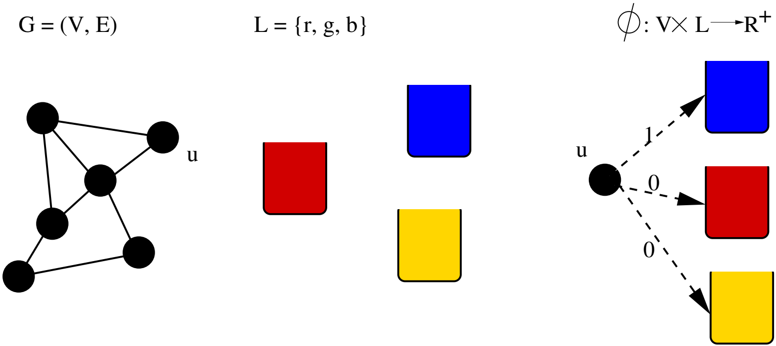
Metric Labeling

$G = (V, E)$ $L = \{r, g, b\}$

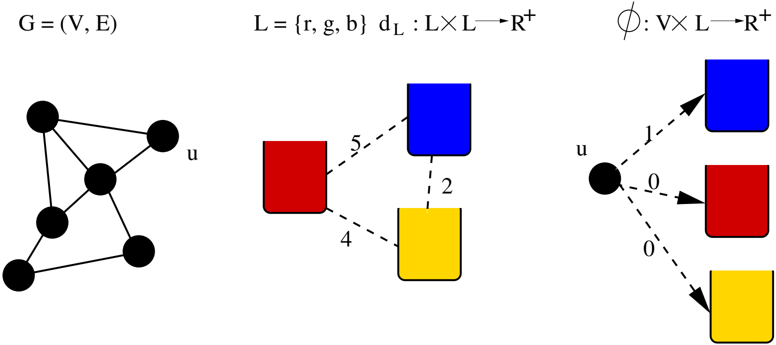




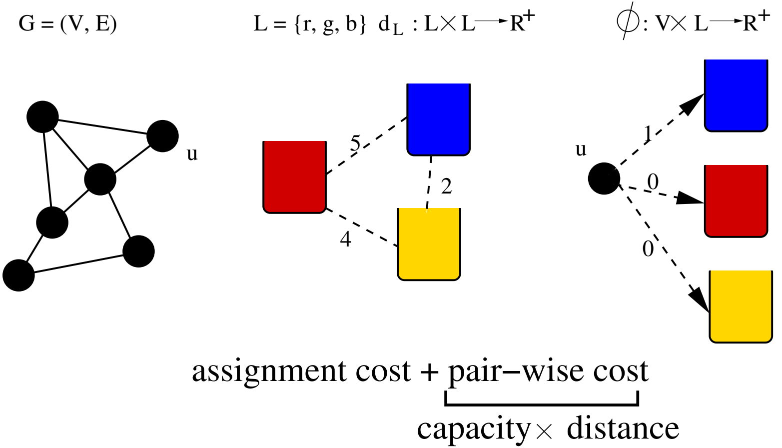
Metric Labeling



Metric Labeling



Metric Labeling



Metric Labeling Problem: (introduced by Kleinberg and Tardos)

$$\min_{f:V\to L} \sum_{u\in V} \phi(u, f(u)) + \sum_{(u,v)\in E} w(u,v) d_L(f(u), f(v))$$

Metric Labeling Problem: (introduced by Kleinberg and Tardos)

$$\min_{f:V\to L} \sum_{u\in V} \phi(u, f(u)) + \sum_{(u,v)\in E} w(u,v) d_L(f(u), f(v))$$

Applications to classification problems in statistical physics, biometry, machine vision, ...

Metric Labeling Problem: (introduced by Kleinberg and Tardos)

$$\min_{f:V\to L}\sum_{u\in V}\phi(u,f(u))+\sum_{(u,v)\in E}w(u,v)d_L(f(u),f(v))$$

Applications to classification problems in statistical physics, biometry, machine vision, ...

Also encodes MAP estimation problem for Markov Random Fields

Let $|V| = n, |L| = k$

Theorem (Kleinberg, Tardos)

There is a polynomial time $O(\log k)$-approximation algorithm for METRIC LABELING

Let $|V| = n, |L| = k$

Theorem (Kleinberg, Tardos)

There is a polynomial time $O(\log k)$-approximation algorithm for METRIC LABELING

Theorem (Chuzhoy, Naor)

For any $\epsilon > 0$, METRIC LABELING *is $\Omega(\log^{1/2-\epsilon} k)$-hard to approximate, unless $P = NP$*

Let $|V| = n, |L| = k$

Theorem (Kleinberg, Tardos)

There is a polynomial time $O(\log k)$-approximation algorithm for Metric Labeling

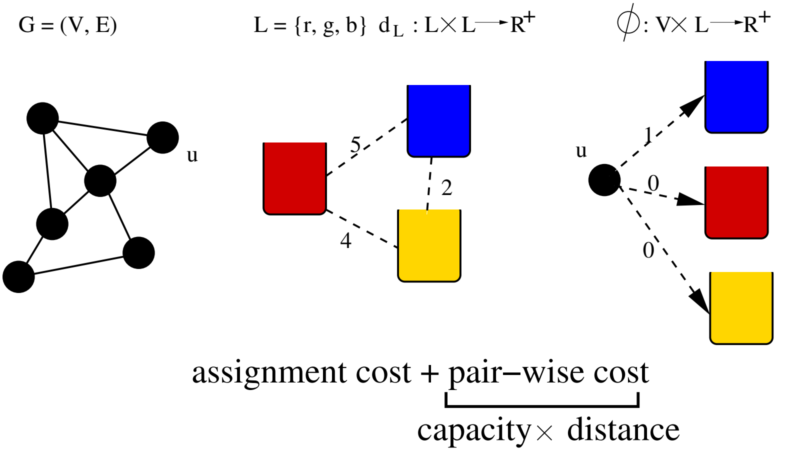
Theorem (Chuzhoy, Naor)

For any $\epsilon > 0$, Metric Labeling *is $\Omega(\log^{1/2-\epsilon} k)$-hard to approximate, unless $P = NP$*

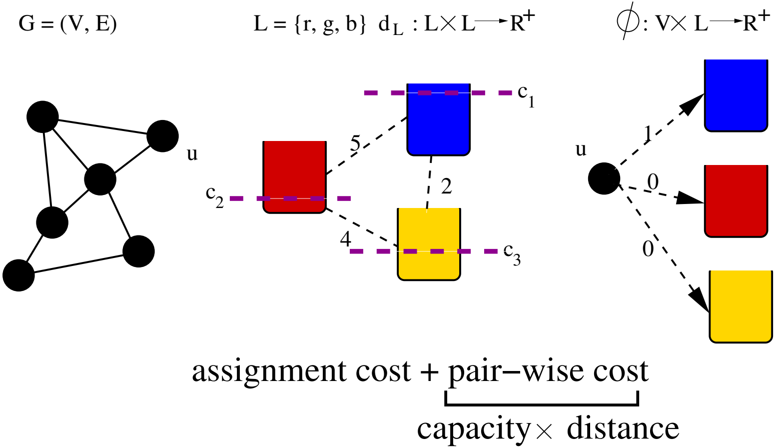
Problem

What if an approximation algorithm returns a highly imbalanced (balanced) solution, and our goal is a balanced (imbalanced) solution?

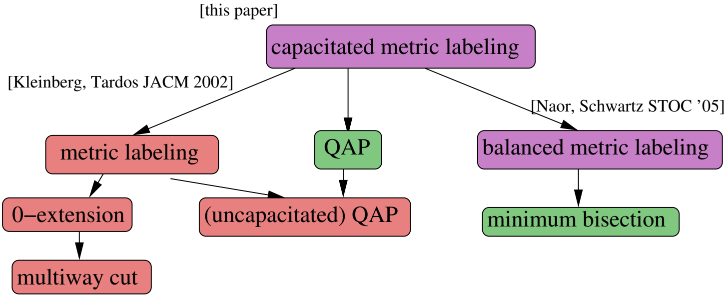
Capacitated Metric Labeling



Capacitated Metric Labeling



Optimization



Our Results

Theorem

For $k = O(1)$, there is a polynomial time $O(\log n)$-approximation algorithm for CAPACITATED METRIC LABELING

Our Results

Theorem

For $k = O(1)$*, there is a polynomial time* $O(\log n)$*-approximation algorithm for* CAPACITATED METRIC LABELING

... this is the regime of interest for many classification problems

Congestion

Question

What if $|L|$ is not constant?

Congestion

Question

What if $|L|$ is not constant?

In this case, determining if there is a ZERO cost solution is HARD

Congestion

Question

What if $|L|$ is not constant?

In this case, determining if there is a ZERO cost solution is HARD

Definition

The congestion an instance I of Capacitated Metric Labeling is the minimum value of C so that scaling the label capacities up by a factor of C has a zero cost solution

Our Results

Theorem

For $k = O(1)$, there is a polynomial time $O(\log n)$-approximation algorithm for CAPACITATED METRIC LABELING

... this is the regime of interest for many classification problems

Our Results

Theorem

For $k = O(1)$, there is a polynomial time $O(\log n)$-approximation algorithm for CAPACITATED METRIC LABELING

... this is the regime of interest for many classification problems

Theorem

For any $\epsilon > 0$, the congestion of CAPACITATED METRIC LABELING *is $\Omega(\log^{1/2-\epsilon} k)$-hard to approximate, unless $NP \subseteq ZPTIME(n^{\text{polylog } n})$*

Our Results

Theorem

For $k = O(1)$, there is a polynomial time $O(\log n)$-approximation algorithm for Capacitated Metric Labeling

... this is the regime of interest for many classification problems

Theorem

For any $\epsilon > 0$, the congestion of Capacitated Metric Labeling *is $\Omega(\log^{1/2-\epsilon} k)$-hard to approximate, unless $NP \subseteq ZPTIME(n^{\text{polylog } n})$*

Theorem

There is a polynomial time $O(\log k)$-approximation algorithm for the congestion of Capacitated Metric Labeling

Our Results

Theorem

For $k = O(1)$, there is a polynomial time $O(\log n)$-approximation algorithm for CAPACITATED METRIC LABELING

... this is the regime of interest for many classification problems

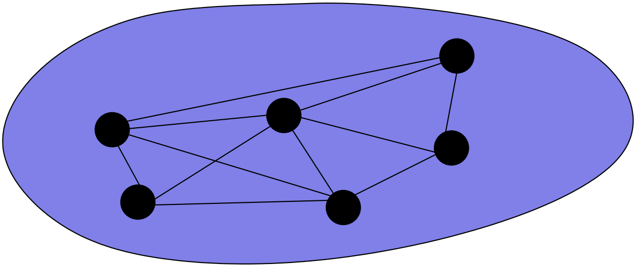
Theorem

For any $\epsilon > 0$, the congestion of CAPACITATED METRIC LABELING *is $\Omega(\log^{1/2-\epsilon} k)$-hard to approximate, unless $NP \subseteq ZPTIME(n^{\text{polylog } n})$*

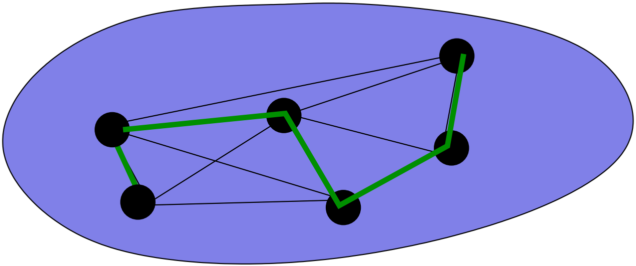
Theorem

There is a polynomial time $O(\log k)$-approximation algorithm for the congestion of CAPACITATED METRIC LABELING

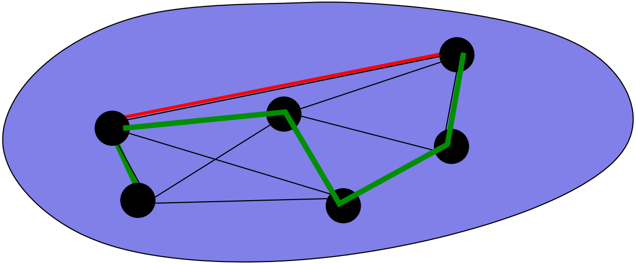
Hierarchical Decompositions



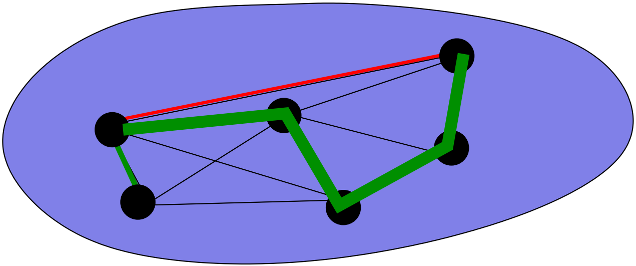
Hierarchical Decompositions



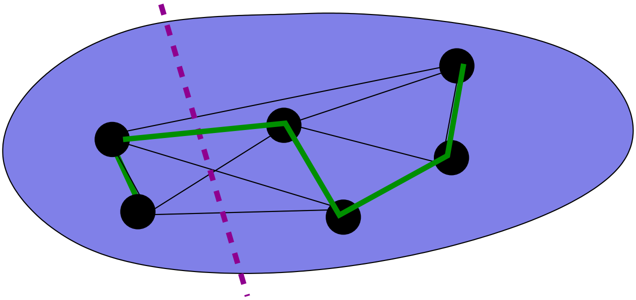
Hierarchical Decompositions



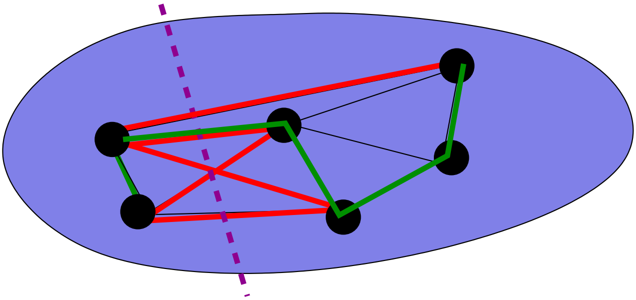
Hierarchical Decompositions



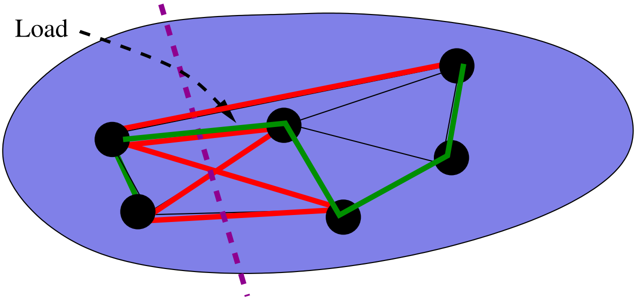
Hierarchical Decompositions



Hierarchical Decompositions



Hierarchical Decompositions

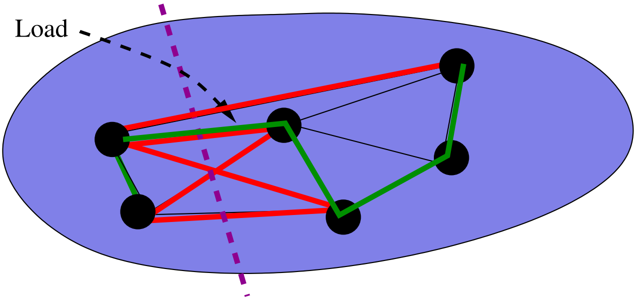


Hierarchical Decompositions

Theorem (Räcke)

There is a distribution μ on decomposition trees so that for all edges,

$$E_{T \leftarrow \mu}[load_T(e)] \leq O(\log n) w(e)$$



Rounding to a Tree

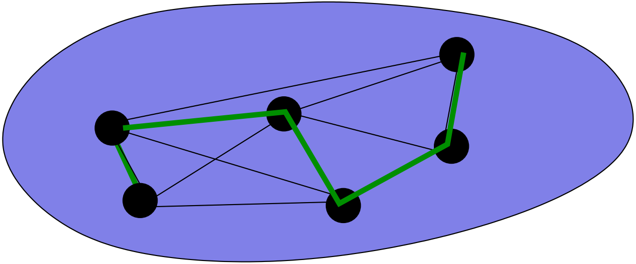
Lemma

$COST(f, G) \leq COST(f, T)$

Rounding to a Tree

Lemma

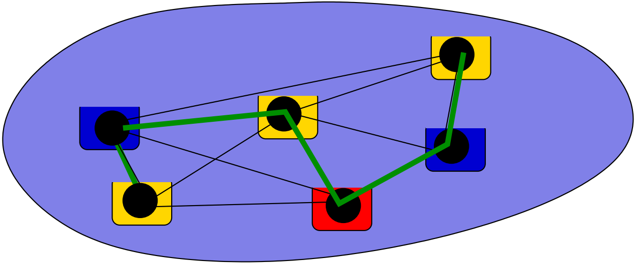
$COST(f, G) \leq COST(f, T)$



Rounding to a Tree

Lemma

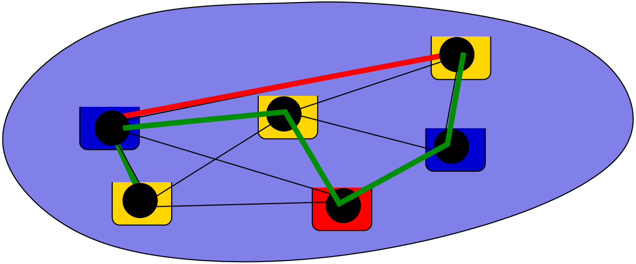
$COST(f, G) \leq COST(f, T)$



Rounding to a Tree

Lemma

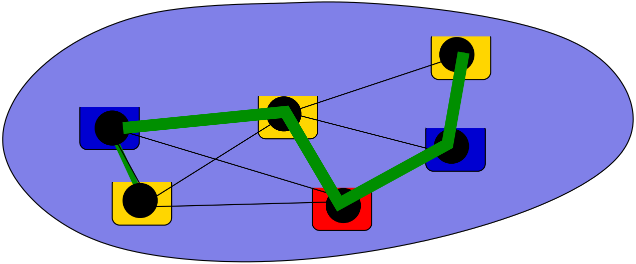
$COST(f, G) \leq COST(f, T)$



Rounding to a Tree

Lemma

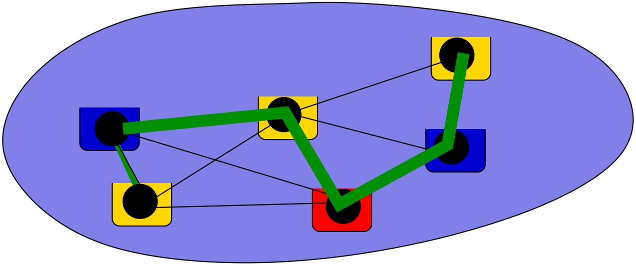
$COST(f, G) \leq COST(f, T)$



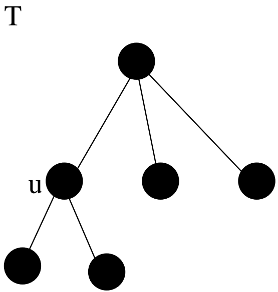
Rounding to a Tree

Lemma

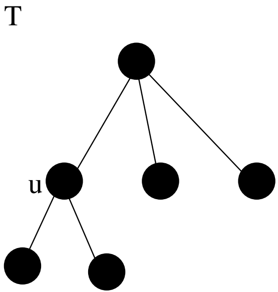
$COST(f,G) \leq COST(f,T)$ *and* $E_{T \leftarrow \mu}[COST(f,T)] \leq O(\log n) COST(f,G)$

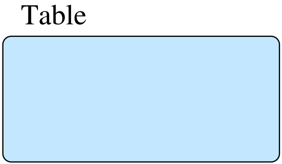


Dynamic Programming

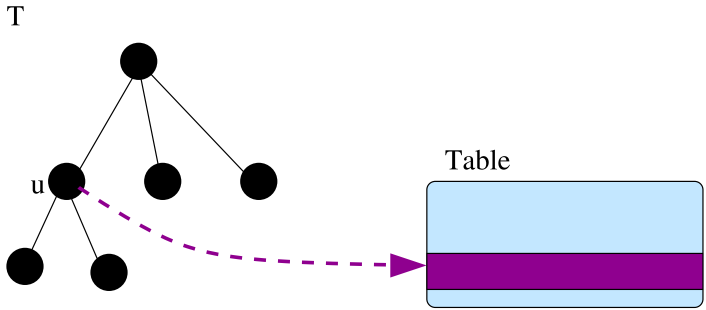


Dynamic Programming

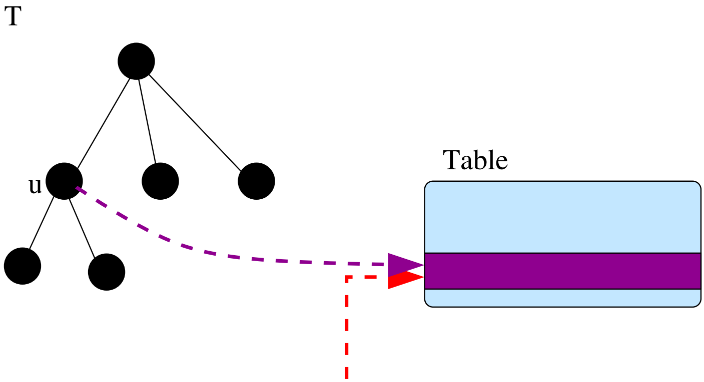




Dynamic Programming



Dynamic Programming



$$\{ \phi(u), c'_1, c'_2, \ldots c'_k \}$$

Our Results

Theorem

For $k = O(1)$, there is a polynomial time $O(\log n)$-approximation algorithm for CAPACITATED METRIC LABELING

... this is the regime of interest for many classification problems

Theorem

For any $\epsilon > 0$, the congestion of CAPACITATED METRIC LABELING *is $\Omega(\log^{1/2-\epsilon} k)$-hard to approximate, unless $NP \subseteq ZPTIME(n^{polylog\ n})$*

Theorem

There is a polynomial time $O(\log k)$-approximation algorithm for the congestion of CAPACITATED METRIC LABELING

Our Results

Theorem

For $k = O(1)$, there is a polynomial time $O(\log n)$-approximation algorithm for CAPACITATED METRIC LABELING

... this is the regime of interest for many classification problems

Theorem

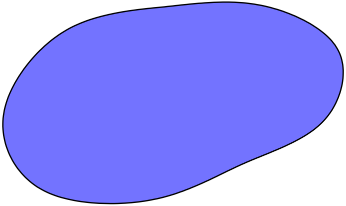
For any $\epsilon > 0$, the congestion of CAPACITATED METRIC LABELING *is $\Omega(\log^{1/2-\epsilon} k)$-hard to approximate, unless $NP \subseteq ZPTIME(n^{\text{polylog } n})$*

Theorem

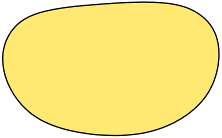
There is a polynomial time $O(\log k)$-approximation algorithm for the congestion of CAPACITATED METRIC LABELING

Approximating Congestion

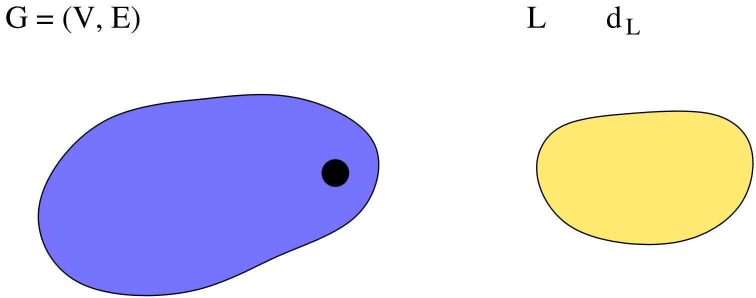
$G = (V, E)$



L d_L

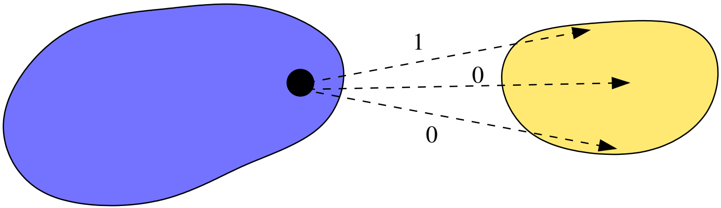


Approximating Congestion

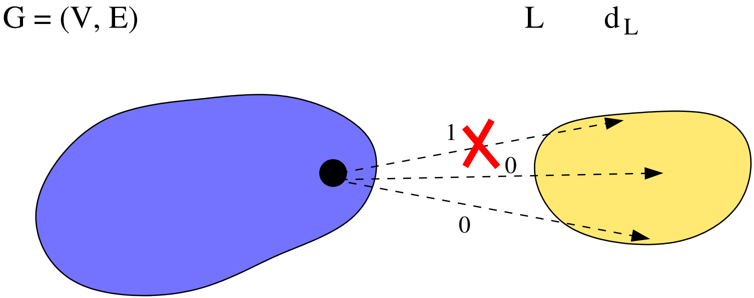


Approximating Congestion

$G = (V, E)$ L d_L

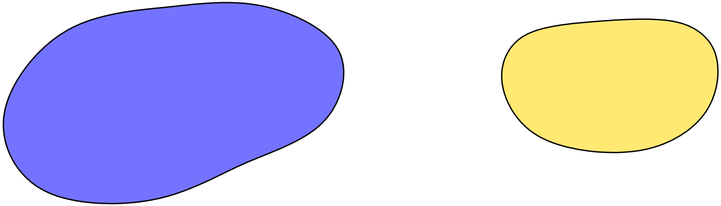


Approximating Congestion

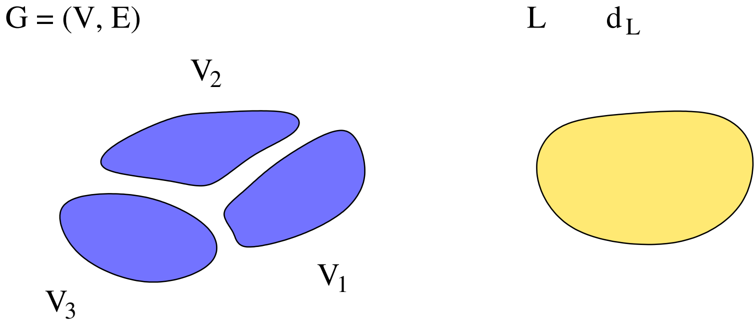


Approximating Congestion

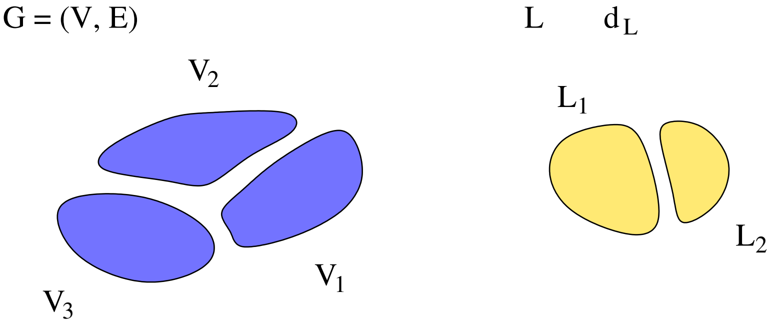
$G = (V, E)$ $\quad$ L $\quad$ d_L



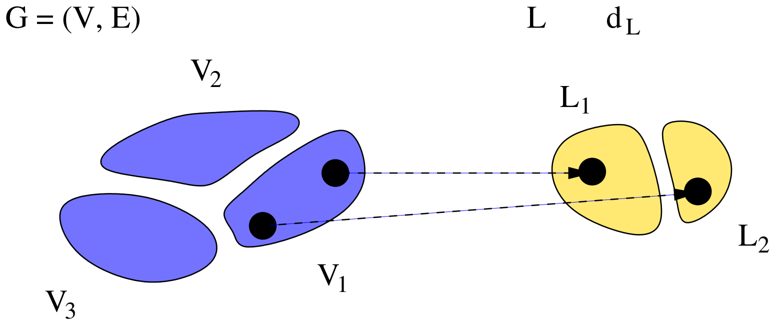
Approximating Congestion



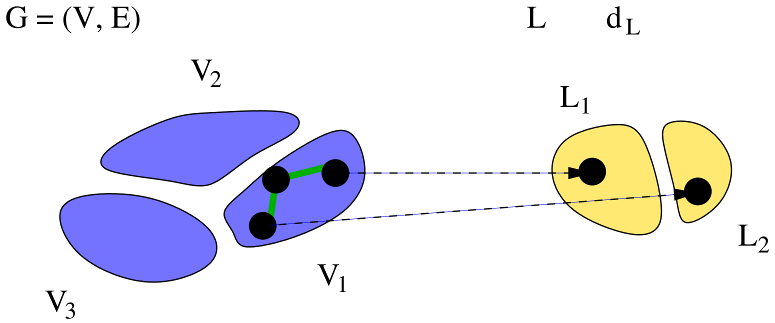
Approximating Congestion



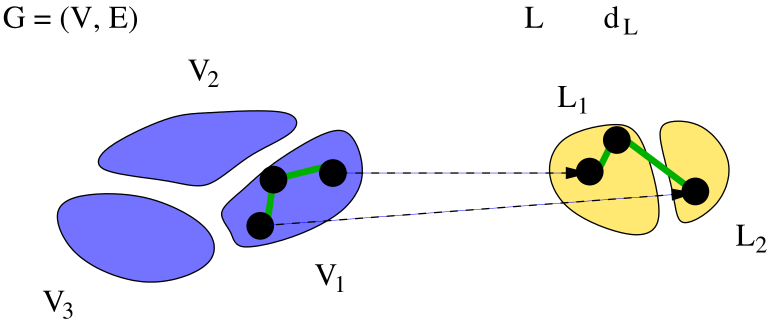
Approximating Congestion



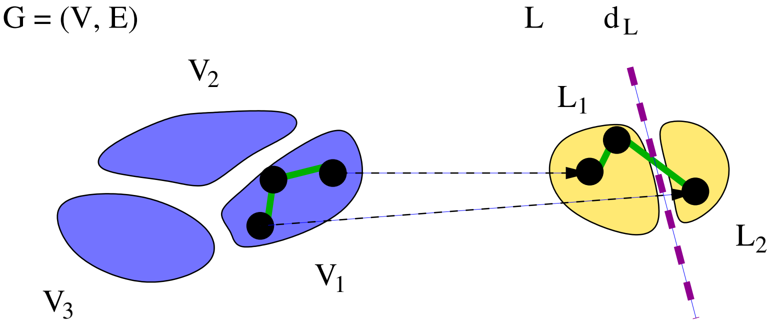
Approximating Congestion



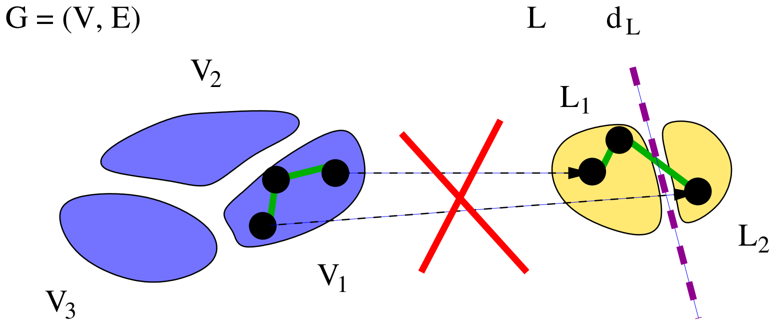
Approximating Congestion



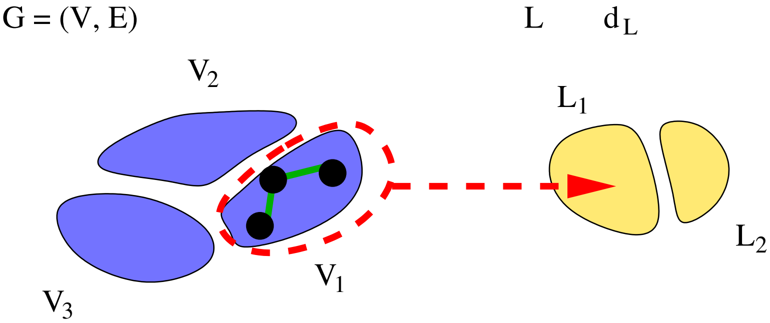
Approximating Congestion



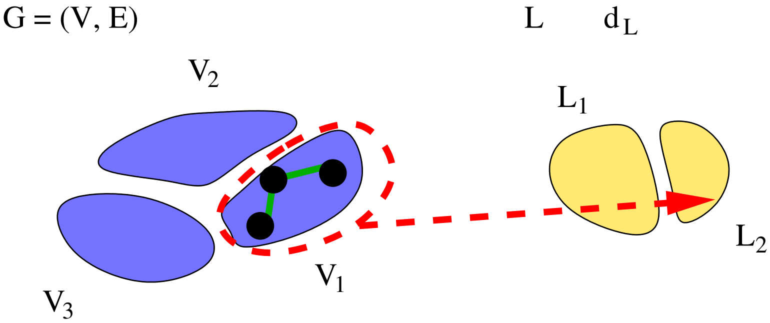
Approximating Congestion



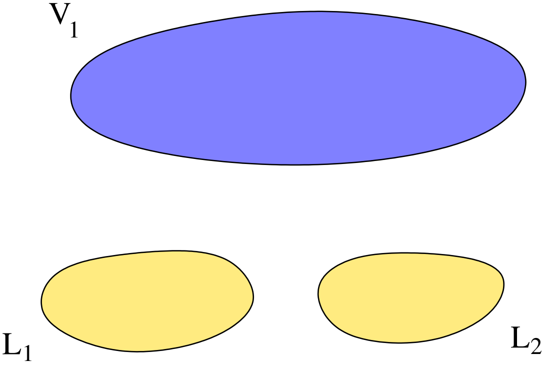
Approximating Congestion



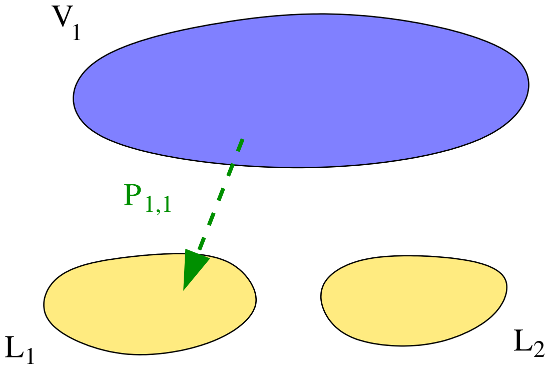
Approximating Congestion



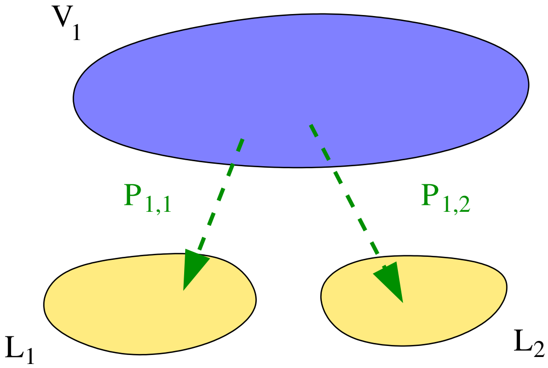
Conditional Probability



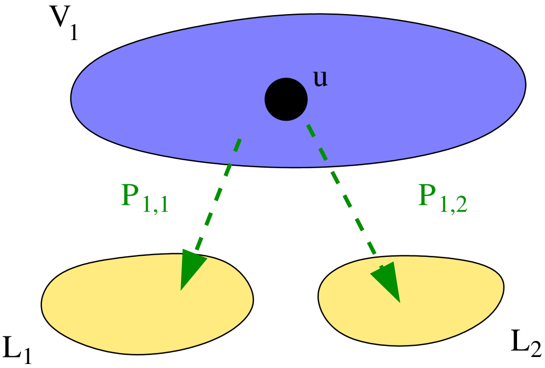
Conditional Probability



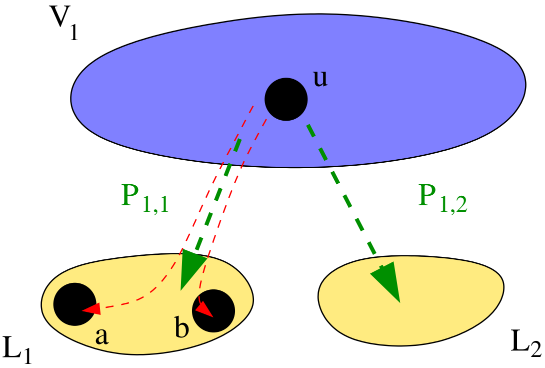
Conditional Probability



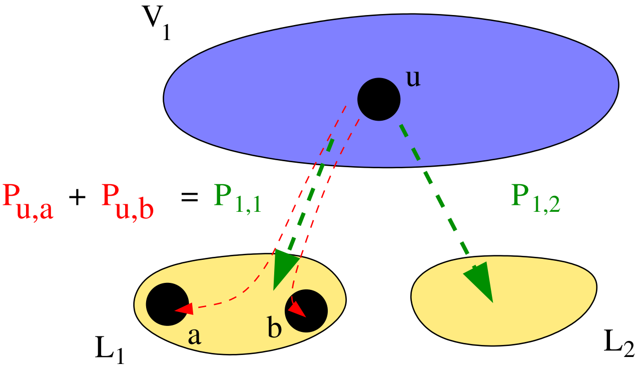
Conditional Probability



Conditional Probability



Conditional Probability



Two Level Rounding

Procedure

Two Level Rounding

Procedure

1. *For each component* V_i*:*

$$\textit{Choose } V_i \rightarrow L_j \textit{ according to } P_{i,j}$$

Two Level Rounding

Procedure

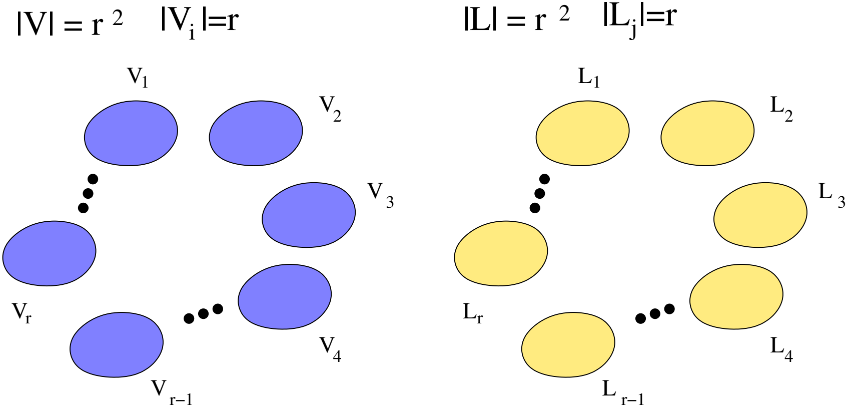
1. *For each component* V_i*:*

   Choose $V_i \rightarrow L_j$ *according to* $P_{i,j}$

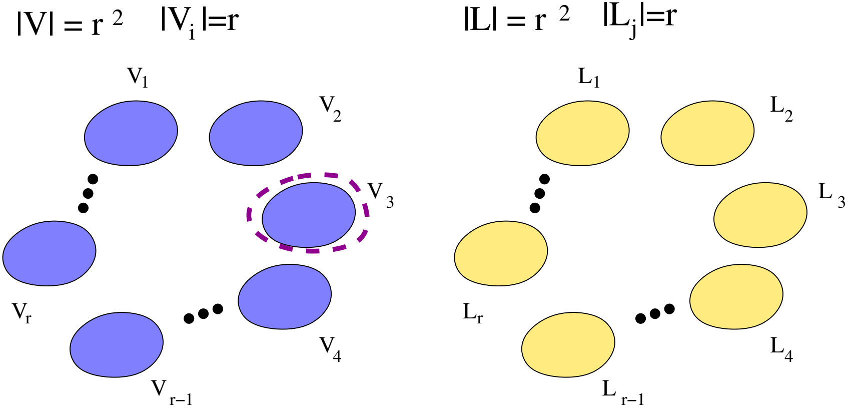
2. *For each component* V_i *mapped to* L_j*, for each* $u \in V_i$*:*

   Choose $u \rightarrow a$ *according to* $\frac{P_{u,a}}{P_{i,j}}$

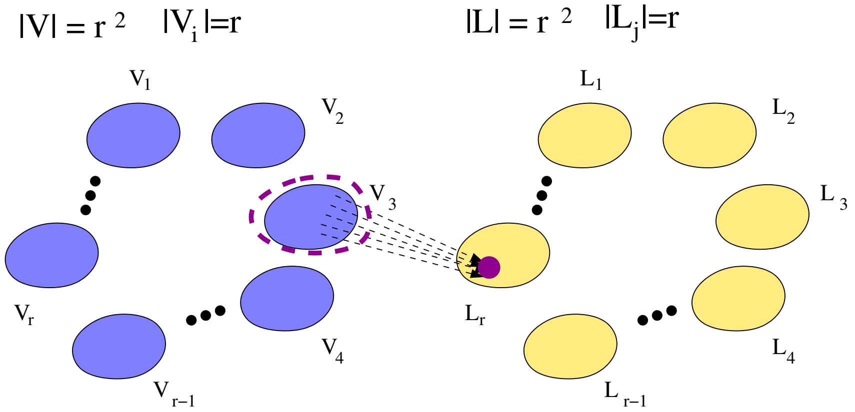
An Integrality Gap



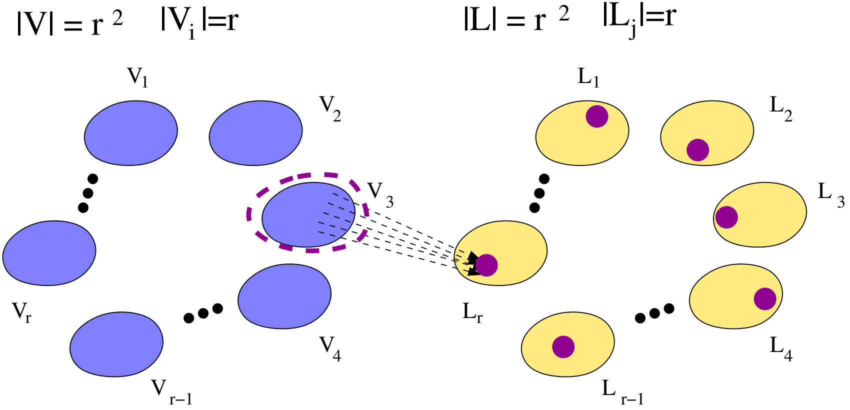
An Integrality Gap



An Integrality Gap



An Integrality Gap



Planning for Rounding

Expectation: $E\left[|\{u \in V_1 | \gamma(u) = a\}|\right] = \sum_{u \in V_1} P_{u,a}$

Planning for Rounding

Expectation: $E\left[|\{u \in V_1 | \gamma(u) = a\}|\right] = \sum_{u \in V_1} P_{u,a}$

Conditional Expectation: $E\left[|\{u \in V_1 | \gamma(u) = a\}| \middle| V_1 \to L_1\right] = \frac{\sum_{u \in V_1} P_{u,a}}{Pr[V_1 \to L_1]}$

Planning for Rounding

Expectation: $E\left[|\{u \in V_1 | \gamma(u) = a\}|\right] = \sum_{u \in V_1} P_{u,a}$

Conditional Expectation: $E\left[|\{u \in V_1 | \gamma(u) = a\}| \middle| V_1 \to L_1\right] = \frac{\sum_{u \in V_1} P_{u,a}}{Pr[V_1 \to L_1]}$

Observation

In an integral solution the conditional expectation (of $a \in L_1$) is also bounded by the label capacity (of a)

Conditional Expectation II

Let $X_1, X_2, \ldots X_T$ be the expectation for a label a

Conditional Expectation II

Let $X_1, X_2, ...X_T$ be the expectation for a label a

Claim

$X_1, X_2, ...X_T$ is a martingale

Conditional Expectation II

Let $X_1, X_2, ...X_T$ be the expectation for a label a

Claim

$X_1, X_2, ...X_T$ is a martingale

Claim

$|X_{i+1} - X_i|$ is at most the capacity of label a

Conditional Expectation II

Let $X_1, X_2, ...X_T$ be the expectation for a label a

Claim

$X_1, X_2, ...X_T$ is a martingale

Claim

$|X_{i+1} - X_i|$ is at most the capacity of label a

Theorem

There is a polynomial time $O(\log k)$-approximation algorithm for the congestion of CAPACITATED METRIC LABELING

Open Questions

Open Question

Is there a polynomial time, poly-logarithmic approximation algorithm for Capacitated Metric Labeling *that violates label capacities multiplicatively by* $O(\log k)$*?*

Open Questions

Open Question

Is there a polynomial time, poly-logarithmic approximation algorithm for Capacitated Metric Labeling *that violates label capacities multiplicatively by $O(\log k)$?*

Analogy: GAP (due to Shmoys-Tardos) compared to scheduling unrelated parallel machines (due to Lenstra-Shmoys-Tardos)

Open Questions

Open Question

Is there a polynomial time, poly-logarithmic approximation algorithm for CAPACITATED METRIC LABELING *that violates label capacities multiplicatively by* $O(\log k)$*?*

Analogy: GAP (due to Shmoys-Tardos) compared to scheduling unrelated parallel machines (due to Lenstra-Shmoys-Tardos)

Open Question

Can the notion of congestion be used to give bi-criteria hardness for other (graph partitioning) problems?

Questions?

Thanks!