Taxicab Diversities

David Bryant (Otago) and Paul Tupper (Simon Fraser)

Overview

1. Taxicab ($L1$) metrics and how they can be used to solve hard problems.
2. The idea of a metric generalizes: introducing the diversity.
3. Harder problems on graphs (and hypergraphs) can be solved using taxicab diversities.

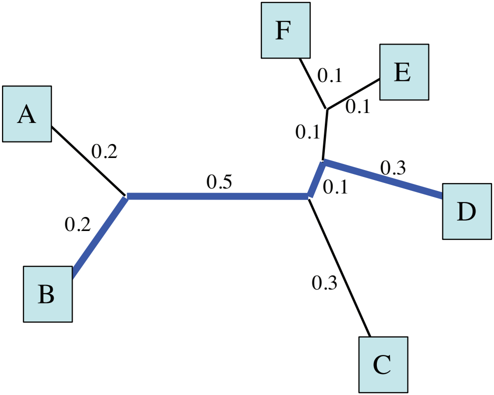
Metrics

A metric on a set satisfies

1. $d(a, b) = d(b, a) \geq 0$ for all a, b.
2. $d(a, b) = 0$ exactly when $a = b$.
3. $d(a, b) \leq d(a, c) + d(b, c)$ for all a, b, c.

The combination of a set with a metric on that set is called a metric space.

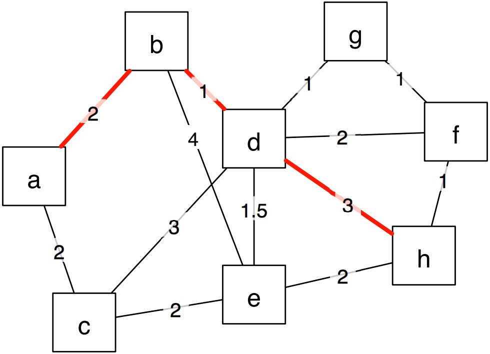
Distances in a tree



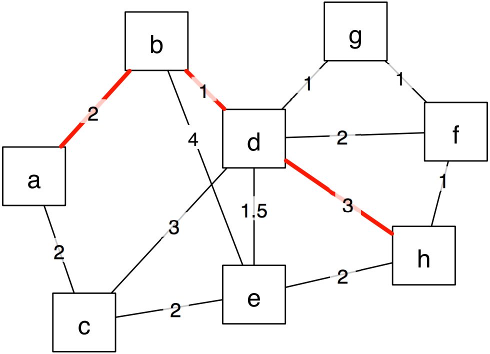
Distance from B to D is the length of the path connecting them:

$$d(B, D) = 0.2 + 0.5 + 0.1 + 0.3.$$

Distances in a graph

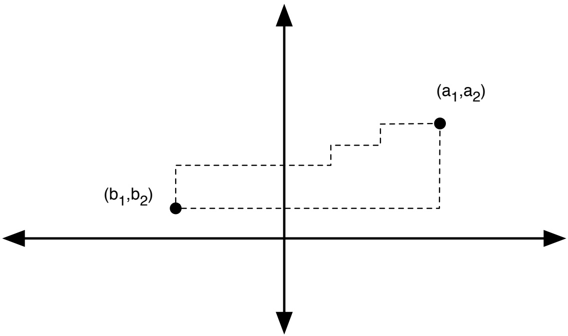


Distances in a graph



This is the maximum metric such that $d(u, v) \leq \ell(u, v)$ for all edges u, v.

Taxicab metric (a.k.a. L_1 or Manhattan metric)



$$d(a,b) = |a_1 - b_1| + |a_2 - b_2|$$

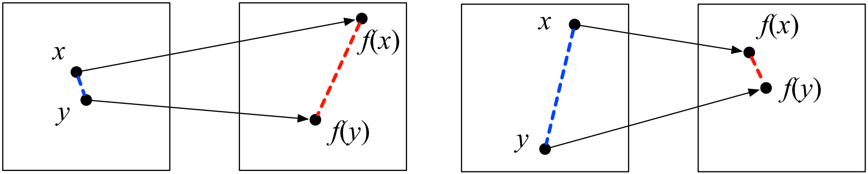
Generalizes to multiple dimensions.

Distortion

Given a function f, how much do distances between points get expanded or shrunk?

One measure is the distortion

$$\left(\max_{x,y} \frac{d_2(f(x), f(y))}{d_1(x,y)}\right) \cdot \left(\max_{x,y} \frac{d_1(x,y)}{d_2(f(x), f(y))}\right).$$



The famous theorems

Johnson-Lindenstrauss Lemma. Any set of m points in high dimensional Euclidean space can be embedded in small $O(\epsilon^{-1} \log m)$ dimensional space with distortion $(1+\epsilon)$.

The famous theorems

Johnson-Lindenstrauss Lemma. Any set of m points in high dimensional Euclidean space can be embedded in small $O(\epsilon^{-1} \log m)$ dimensional space with distortion $(1+\epsilon)$.

Bourgain's theorem. Any metric on n points can be embedded in $\log^2 n$ dimensional L_1 space with distortion at most $\log n$.

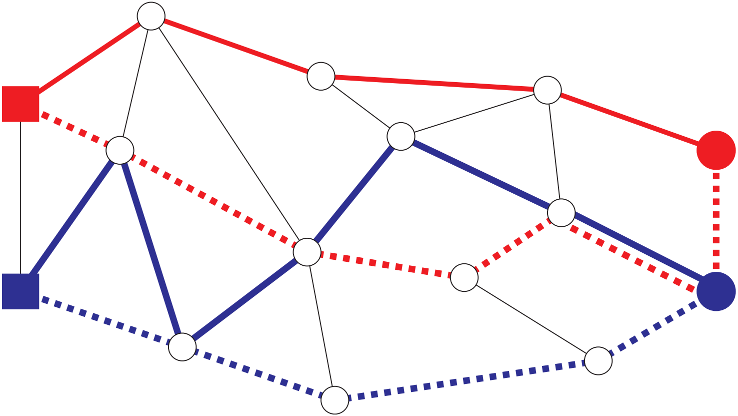
The famous theorems

Johnson-Lindenstrauss Lemma. Any set of m points in high dimensional Euclidean space can be embedded in small $O(\epsilon^{-1} \log m)$ dimensional space with distortion $(1 + \epsilon)$.

Bourgain's theorem. Any metric on n points can be embedded in $\log^2 n$ dimensional L_1 space with distortion at most $\log n$.

Applications in large scale clustering, pattern matching, large data. The use of small distortion mappings has been one of the big ideas in algorithm design over the past 10-15 years.

Flow and cut



Flow and cut

Multi-commodity flow

Input: Demands D_{uv} and edge capacities C_{uv}.

Problem: Maximize λ such that we can simultaneously flow λD_{uv} between all u, v.

Sparsest Cut

Input: Demands D_{uv} and edge capacities C_{uv}

Problem: Find a cut $U|V$ which minimizes

$$\frac{\sum_{u \in U, v \in V} C_{uv}}{\sum_{u \in U, v \in V} D_{uv}}$$

Flow and cut

Multi-commodity flow

Input: Demands D_{uv} and edge capacities C_{uv}.

Problem: Maximize λ such that we can simultaneously flow λD_{uv} between all u, v.

Sparsest Cut

Input: Demands D_{uv} and edge capacities C_{uv}

Problem: Find a cut $U|V$ which minimizes

$$\frac{\sum_{u \in U, v \in V} C_{uv}}{\sum_{u \in U, v \in V} D_{uv}}$$

The maximum flow is always less than or equal to size of the sparsest cut.

Sparsest cut via L_1 embedding

It can be shown (LP duality) that multicommodity flow is equivalent to

$$\min \sum_{uv} C_{uv} d(u,v)$$

such that $\sum_{uv} D_{uv} d(u,v) \geq 1$ and d is a metric.

Sparsest cut via L_1 embedding

It can be shown (LP duality) that multicommodity flow is equivalent to

$$\min \sum_{uv} C_{uv} d(u,v)$$

such that $\sum_{uv} D_{uv} d(u,v) \geq 1$ and d is a metric.

It can also be shown that sparsest cut is equivalent to

$$\min \sum_{uv} C_{uv} d(u,v)$$

such that $\sum_{uv} D_{uv} d(u,v) \geq 1$ and d is an L_1 metric*

Approximating Sparsest Cut

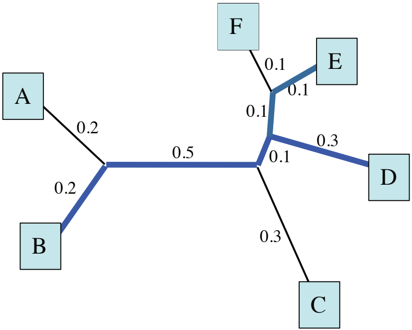
1. Solve the dual of multicommodity flow.
2. Find a low distortion embedding of the output of 1. into L_1.
3. Extract a solution to sparsest cut.

From Bourgain's result we obtain an $O(\log n)$ approximation.

Generalizing metrics

What if we go from pairs to triples, 4-sets, finite subsets?

Token phylogenetics



Diversity of B,D,E is the length of the tree connecting them.

$$\delta(\{B, D, E\}) = 0.2 + 0.5 + 0.1 + 0.3 + 0.1 + 0.1$$

Formalising the idea of diversities

Set X of points and a function δ on finite subsets of X.

1. For all A we have $\delta(A) \geq 0$.
2. For all A we have $\delta(A) = 0$ exactly when $|A| \leq 1$.
3. For all A, B, C with $C \neq \emptyset$ we have

$$\delta(A \cup B) \leq \delta(A \cup C) + \delta(C \cup B).$$

A pair (X, δ) satisfying all of these is called a diversity. (First presented at Phylomania '09)

Formalising the idea of diversities

Set X of points and a function δ on finite subsets of X.

1. For all A we have $\delta(A) \geq 0$.
2. For all A we have $\delta(A) = 0$ exactly when $|A| \leq 1$.
3. For all A, B, C with $C \neq \emptyset$ we have

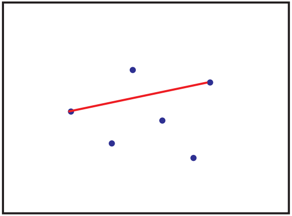
$$\delta(A \cup B) \leq \delta(A \cup C) + \delta(C \cup B).$$

A pair (X, δ) satisfying all of these is called a diversity. (First presented at Phylomania '09)

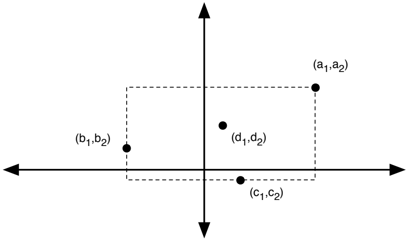
Note that δ restricted to pairs is a metric.

Examples: diameter diversity

Let (X, d) be a metric space. Define $\delta(A) = \max_{a,b \in A} d(a, b)$. Then (X, δ) is a diversity.



Example: Taxicab (L_1 or Manhattan) diversities

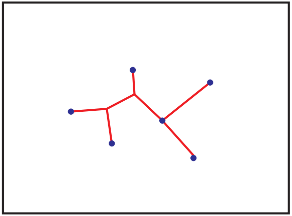


Diversity of a set of points is the height+width of the smallest box containing them.

$$\delta(\{a, b, c\}) = |a_1 - b_1| + |a_2 - c_2|$$

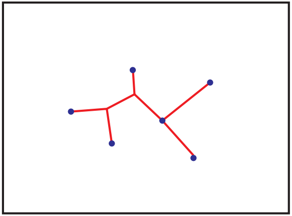
Example: Steiner tree

Let (X, d) be a metric space. For each finite $A \subseteq X$ let $\delta(A)$ be the length of the minimum Steiner tree connecting A. Then (X, δ) is a diversity.



Example: Steiner tree

Let (X, d) be a metric space. For each finite $A \subseteq X$ let $\delta(A)$ be the length of the minimum Steiner tree connecting A. Then (X, δ) is a diversity.



On graphs, the Steiner tree diversity is to diversities what the shortest path metric is to metrics.

Diversity theory

We've now got many different examples of diversities, from TSP to Steiner trees to geometric probability. **Introduction, tight spans and hyperconvexity** Bryant & Tupper (2012) *Advances Math.*
Results on L_1 diversities Bryant & Klaere (2011) *J. Math. Bio.*
Polyhedral Formulation of Diversity Tight Span: Herrmann & Moulton (2012) *Discrete Math.*
Connections to Order Theory: Ben Whale (2013).
Fixed Point Theory: Piatek & Espinola (2013).
Analogue of Uniform Spaces: Poelstra (2013).
Geometry of hypergraphs Bryant & Tupper (2014)
More fixed point theory: Kirk & Shahzad (2014)

Geometry of graphs revisited

Let $G = (V, E)$ be a graph with edge weights. The *shortest path metric* is the maximal metric such that

$$\ell(u, v) \geq d(u, v)$$

for all edges $\{u, v\}$. What is the diversity analogue?

Geometry of graphs revisited

Let $G = (V, E)$ be a graph with edge weights. The *shortest path metric* is the maximal metric such that

$$\ell(u, v) \geq d(u, v)$$

for all edges $\{u, v\}$. What is the diversity analogue?

The maximal diversity such that $\delta(\{u, v\}) \leq \ell(u, v)$ for all edges $\{u, v\}$ is the *Steiner tree diversity*:

$$\delta(A) = \text{ length of min. Steiner tree connecting } A.$$

What is the *Geometry of Graphs* for diversities?

Distortion and diversities

Johnson-Lindenstrauss Lemma Revisited. Any set of m points in high dimensional Euclidean diversity can be embedded in small $O(\epsilon^{-1} \log m)$ dimensional space with distortion $(1 + \epsilon)$.

Distortion and diversities

Johnson-Lindenstrauss Lemma Revisited. Any set of m points in high dimensional Euclidean diversity can be embedded in small $O(\epsilon^{-1}\log m)$ dimensional space with distortion $(1+\epsilon)$.

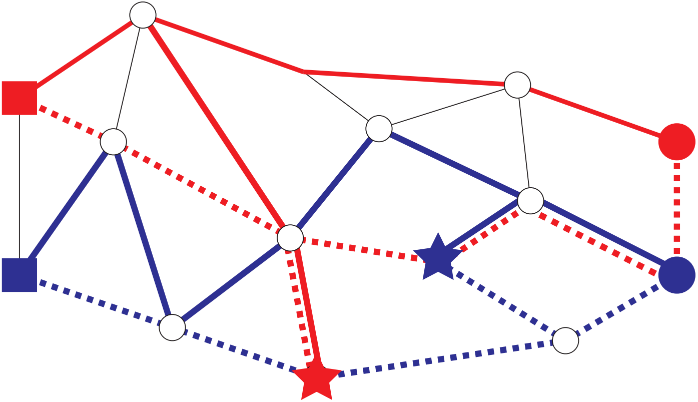
Bourgain's Theorem Revisited. Any diversity on n points can be embedded in $\log^2 n$ dimensional L_1 space with distortion at most $n(\log n)^2$. Conjecture: this should be $O(\log n)$.

Multicommodity Steiner flows

A flow from s_1 to s_2 can be written as the sum of paths from s_1 to s_2. Flow is conserved at nodes.

A *Steiner flow* for S can be written as the sum of trees connecting nodes in S. This models flow of information with broadcasting.

Steiner flows



Flow and cut 2

Multi-commodity (concurrent) Steiner Flow

Input: Demands D_S and edge capacities C_{uv}.

Problem: Maximize λ such that we can simultaneously share signal between S at rate λD_S.

Sparsest Cut take 2

Input: Demands D_S for $S \subseteq V$ and edge capacities C_{uv}

Problem: Find a cut $U|V$ which minimizes

$$\frac{\sum_{u \in U, v \in V} C_{uv}}{\sum_{\substack{S \cap U \neq \emptyset \\ S \cap V \neq \emptyset}} D_S}$$

Flow and cut 2

Multi-commodity (concurrent) Steiner Flow

Input: Demands D_S and edge capacities C_{uv}.

Problem: Maximize λ such that we can simultaneously share signal between S at rate λD_S.

Sparsest Cut take 2

Input: Demands D_S for $S \subseteq V$ and edge capacities C_{uv}

Problem: Find a cut $U|V$ which minimizes

$$\frac{\sum_{u \in U, v \in V} C_{uv}}{\sum_{\substack{S \cap U \neq \emptyset \\ S \cap V \neq \emptyset}} D_S}$$

Sparsest cut via L_1 embedding

It can be shown (LP duality) that multicommodity (concurrent) Steiner Flow is equivalent to

$$\min \sum_{uv} C_{uv} \delta(\{u, v\})$$

such that $\sum_S D_S \delta(S) \geq 1$ and δ is a diversity.

Sparsest cut via L_1 embedding

It can be shown (LP duality) that multicommodity (concurrent) Steiner Flow is equivalent to

$$\min \sum_{uv} C_{uv} \delta(\{u, v\})$$

such that $\sum_S D_S \delta(S) \geq 1$ and δ is a diversity.

It can also be shown that the generalized sparsest cut is equivalent to

$$\min \sum_{uv} C_{uv} \delta(\{u, v\})$$

such that $\sum_S D_S \delta(S) \geq 1$ and δ is an L_1 diversity.

Embedding

Theorem Bryant and Tupper:

The Steiner tree diversity for any graph can be embedded in L_1 with distortion $O(\log n)$.

Hence we obtain a generalization of Linial et al's result.

Embedding

Theorem Bryant and Tupper:

The Steiner tree diversity for any graph can be embedded in L_1 with distortion $O(\log n)$.

Hence we obtain a generalization of Linial et al's result.
(see also Klein et al 1997)

All of this can be done for hypergraphs too (although the distortion required is still unknown).

Portobello, 1-6 February, 2015

Bryant, D. and Tupper, P., (2012) Hyperconvexity and tight span theory for diversities. *Advances in Mathematics* 231:3172-3198

Bryant, D. and Tupper, P. (2014) The geometry of hypergraphs. *Discrete Math and Computer Science*