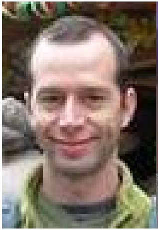
Random points in the metric polytope

Gady Kozma (speaker), Tom Meyerovitch, Ron Peled and Wojciech Samotij


Random Combinatorial Structures and Statistical Mechanics, Venice, 2013

Definition

Let X be a metric space with n points.

Definition

Let X be a metric space with n points. We think about the distances as if they were a vector in $\mathbb{R}^{\binom{n}{2}}$.

Definition

Let X be a metric space with n points. We think about the distances as if they were a vector in $\mathbb{R}^{\binom{n}{2}}$. The union of all such points in $\mathbb{R}^{\binom{n}{2}}$ for all finite metric spaces is a *convex cone* (defined by the equations $d_{ij} \leq d_{ik} + d_{jk}$). This cone is called the *metric cone*.

Definition

Let X be a metric space with n points. We think about the distances as if they were a vector in $\mathbb{R}^{\binom{n}{2}}$. The union of all such points in $\mathbb{R}^{\binom{n}{2}}$ for all finite metric spaces is a *convex cone* (defined by the equations $d_{ij} \le d_{ik} + d_{jk}$). This cone is called the *metric cone.*

Since one cannot take a random point in an infinite set, we need to cut it somehow, and we restrict all distances to be ≤ 2 (in other words, we intersect with the cube $[0, 2]^{\binom{n}{2}}$).

Definition

Let X be a metric space with n points. We think about the distances as if they were a vector in $\mathbb{R}^{\binom{n}{2}}$. The union of all such points in $\mathbb{R}^{\binom{n}{2}}$ for all finite metric spaces is a *convex cone* (defined by the equations $d_{ij} \leq d_{ik} + d_{jk}$). This cone is called the *metric cone*.

Since one cannot take a random point in an infinite set, we need to cut it somehow, and we restrict all distances to be ≤ 2 (in other words, we intersect with the cube $[0, 2]^{\binom{n}{2}}$). We will call this set "the metric polytope" even though the standard definition is different.

Definition

Let X be a metric space with n points. We think about the distances as if they were a vector in $\mathbb{R}^{\binom{n}{2}}$. The union of all such points in $\mathbb{R}^{\binom{n}{2}}$ for all finite metric spaces is a *convex cone* (defined by the equations $d_{ij} \leq d_{ik} + d_{jk}$). This cone is called the *metric cone*.

Since one cannot take a random point in an infinite set, we need to cut it somehow, and we restrict all distances to be ≤ 2 (in other words, we intersect with the cube $[0, 2]^{\binom{n}{2}}$). We will call this set "the metric polytope" even though the standard definition is different. How does a random point look like?

The box

We note that the box $[1,2]^{\binom{n}{2}}$ is completely contained in the metric polytope M_n: the triangle condition is always satisfied. Hence $\operatorname{vol} M_n \geq 1$.

The box

We note that the box $[1,2]^{\binom{n}{2}}$ is completely contained in the metric polytope M_n: the triangle condition is always satisfied. Hence $\operatorname{vol} M_n \geq 1$.

Of course, this holds for any box $[a,2a]^{\binom{n}{2}}$, but the contribution of $a < 1$ is negligible.

The box

We note that the box $[1,2]^{\binom{n}{2}}$ is completely contained in the metric polytope M_n: the triangle condition is always satisfied. Hence $\operatorname{vol} M_n \geq 1$.

Of course, this holds for any box $[a,2a]^{\binom{n}{2}}$, but the contribution of $a < 1$ is negligible.

Our main result is that a random point in the metric polytope looks like a random point in this cube.

Theorem

$$1 + \frac{c}{\sqrt{n}} \le (\operatorname{vol} M_n)^{1/\binom{n}{2}} \le 1 + \frac{C}{n^c}.$$

Further,

$$\mathbb{P}(d_{ij} > 1 - Cn^{-c} \quad \forall 1 \le i < j \le n) \ge 1 - Ce^{-cn}.$$

Theorem

$$1 + \frac{c}{\sqrt{n}} \le (\mathrm{vol}\, M_n)^{1/\binom{n}{2}} \le 1 + \frac{C}{n^c}.$$

Further,

$$\mathbb{P}(d_{ij} > 1 - Cn^{-c} \quad \forall 1 \le i < j \le n) \ge 1 - Ce^{-cn}.$$

- In other words, this is useless as a model for a random metric space: the result is too boring.

Theorem

$$1 + \frac{c}{\sqrt{n}} \leq (\mathrm{vol}\, M_n)^{1/\binom{n}{2}} \leq 1 + \frac{C}{n^c}.$$

Further,

$$\mathbb{P}(d_{ij} > 1 - Cn^{-c} \quad \forall 1 \leq i < j \leq n) \geq 1 - Ce^{-cn}.$$

- In other words, this is useless as a model for a random metric space: the result is too boring.
- We will concentrate on the upper bound for the volume, where **entropy** techniques appear.

Theorem

$$1+\frac{c}{\sqrt{n}} \le (\mathrm{vol}\, M_n)^{1/\binom{n}{2}} \le 1+\frac{C}{n^c}.$$

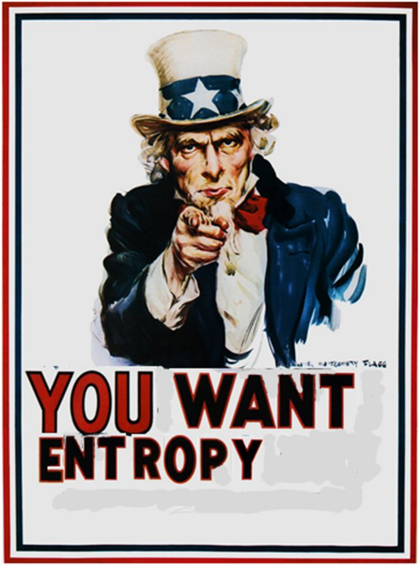
Further,

$$\mathbb{P}(d_{ij} > 1 - Cn^{-c} \quad \forall 1 \le i < j \le n) \ge 1 - Ce^{-cn}.$$

- In other words, this is useless as a model for a random metric space: the result is too boring.
- We will concentrate on the upper bound for the volume, where **entropy** techniques appear. One can get

  $$\mathbb{P}(d_{ij} > \tfrac{1}{32}\ \forall i,j) > 1 - Ce^{-cn}$$

  without going through volume estimates or using entropy (but not trivially).



Graphic design: Ori Kozma

I am sure you all know this, but...

For a variable X taking values in $\mathbb{R}^m$ with density p

$$H(X) = -\int p(x)\log p(x)\,dx$$

(where we define $0\log 0 = 0$). For two variables X and Y,

$$H(X \mid Y) = \mathbb{E}_y(H(X|Y=y)) = H(X,Y) - H(Y).$$

- Like in the discrete case, for a given support the entropy is maximized on the uniform measure.
- Unlike in the discrete case, the entropy can be negative.

Shearer's inequality

Theorem

Let $A_1, \ldots, A_k$ be sets of indices (i.e. each $A_i \subset \{1, \ldots, m\}$) and suppose they r-cover $\{1, \ldots, m\}$ i.e.

$$|\{i : j \in A_i\}| \geq r \qquad \forall j \in \{1, \ldots, m\}$$

Shearer's inequality

Theorem

Let $A_1, \ldots, A_k$ *be sets of indices (i.e. each* $A_i \subset \{1, \ldots, m\}$*) and suppose they* r*-cover* $\{1, \ldots, m\}$ *i.e.*

$$|\{i : j \in A_i\}| \geq r \qquad \forall j \in \{1, \ldots, m\}$$

Let $X_1, \ldots, X_m$ *be (dependent) random variables.*

Shearer's inequality

Theorem

Let $A_1, \ldots, A_k$ be sets of indices (i.e. each $A_i \subset \{1, \ldots, m\}$) and suppose they r-cover $\{1, \ldots, m\}$ i.e.

$$|\{i : j \in A_i\}| \geq r \qquad \forall j \in \{1, \ldots, m\}$$

Let $X_1, \ldots, X_m$ be (dependent) random variables. Then

$$H(X_1, \ldots, X_m) \leq \frac{1}{r} \sum_{i=1}^{k} H(\{X_a\}_{a \in A_i}).$$

(if you prefer: an inequality concerning a measure on $\mathbb{R}^m$ and a collection of its projections)

Shearer's inequality — an application

$$H(X_1, \ldots, X_m) \leq \frac{1}{r} \sum_{i=1}^{k} H(\{x_a\}_{a \in A_i}).$$

Recall that $m = \binom{n}{2}$ so we think about the indices as couples (a, b) with $1 \leq a < b \leq n$.

Shearer's inequality — an application

$$H(X_1, \ldots, X_m) \le \frac{1}{r} \sum_{i=1}^{k} H(\{x_a\}_{a \in A_i}).$$

Recall that $m = \binom{n}{2}$ so we think about the indices as couples (a, b) with $1 \le a < b \le n$. Let the sets A_i correspond to removing one *vertex* (which removes n edges):

$$A_i = \{(a, b) : a \ne i \text{ and } b \ne i\}$$

Shearer's inequality — an application

$$H(X_1, \ldots, X_m) \leq \frac{1}{r} \sum_{i=1}^{k} H(\{x_a\}_{a \in A_i}).$$

Recall that $m = \binom{n}{2}$ so we think about the indices as couples (a, b) with $1 \leq a < b \leq n$. Let the sets A_i correspond to removing one *vertex* (which removes n edges):

$$A_i = \{(a, b) : a \neq i \text{ and } b \neq i\}$$

Our r is now $n - 2$ because each *edge* appears in exactly $n - 2$ sets A_i namely those that do not correspond to either end vertex.

Shearer's inequality — an application

$$H(X_1, \ldots, X_m) \leq \frac{1}{r} \sum_{i=1}^{k} H(\{x_a\}_{a \in A_i}).$$

Recall that $m = \binom{n}{2}$ so we think about the indices as couples (a, b) with $1 \leq a < b \leq n$. Let the sets A_i correspond to removing one *vertex* (which removes n edges):

$$A_i = \{(a, b) : a \neq i \text{ and } b \neq i\}$$

Our r is now $n - 2$ because each *edge* appears in exactly $n - 2$ sets A_i namely those that do not correspond to either end vertex. So

$$\begin{aligned} \log \operatorname{vol} M_n = H(\{d_{ij}\}) &\leq \frac{1}{n-2} \sum_{k=1}^{n} H(\{d_{ij}\}_{i,j \neq k}) \\ &= \frac{n}{n-2} H(\{d_{ij}\}_{i,j \neq n}) \end{aligned}$$

Shearer's inequality — an application

$$\begin{aligned}\log \operatorname{vol} M_n = H(\{d_{ij}\}) &\leq \frac{1}{n-2}\sum_{k=1}^{n} H(\{d_{ij}\}_{i,j\neq k}) \\ &= \frac{n}{n-2} H(\{d_{ij}\}_{i,j\neq n})\end{aligned}$$

Shearer's inequality — an application

$$\begin{aligned}\log \operatorname{vol} M_n = H(\{d_{ij}\}) &\le \frac{1}{n-2}\sum_{k=1}^{n} H(\{d_{ij}\}_{i,j\neq k})\\ &= \frac{n}{n-2} H(\{d_{ij}\}_{i,j\neq n})\end{aligned}$$

But $\{d_{ij}\}_{i,j\neq n}$ is supported on M_{n-1} and thus has entropy smaller than the uniform measure on M_{n-1}.

Shearer's inequality — an application

$$\begin{aligned}\log \operatorname{vol} M_n = H(\{d_{ij}\}) &\leq \frac{1}{n-2}\sum_{k=1}^{n} H(\{d_{ij}\}_{i,j\neq k}) \\ &= \frac{n}{n-2} H(\{d_{ij}\}_{i,j\neq n}) \leq \frac{n}{n-2}\log \operatorname{vol} M_{n-1}\end{aligned}$$

But $\{d_{ij}\}_{i,j\neq n}$ is supported on M_{n-1} and thus has entropy smaller than the uniform measure on M_{n-1}.

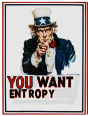
Shearer's inequality — an application

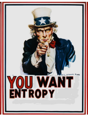
$$\begin{aligned}\log \operatorname{vol} M_n = H(\{d_{ij}\}) &\leq \frac{1}{n-2}\sum_{k=1}^{n} H(\{d_{ij}\}_{i,j\neq k}) \\ &= \frac{n}{n-2} H(\{d_{ij}\}_{i,j\neq n}) \leq \frac{n}{n-2}\log \operatorname{vol} M_{n-1}\end{aligned}$$

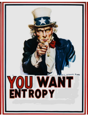
But $\{d_{ij}\}_{i,j\neq n}$ is supported on M_{n-1} and thus has entropy smaller than the uniform measure on M_{n-1}. Rearranging gives:

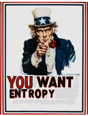
$$\frac{1}{\binom{n}{2}}\log \operatorname{vol} M_n \leq \frac{1}{\binom{n}{2}}\frac{n}{n-2}\log \operatorname{vol} M_{n-1} = \frac{1}{\binom{n-1}{2}}\log \operatorname{vol} M_{n-1}.$$

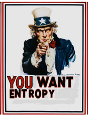
So $\operatorname{vol}(M_n)^{1/\binom{n}{2}}$ is decreasing.

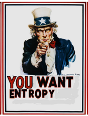


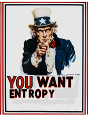


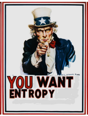


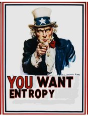


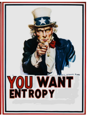






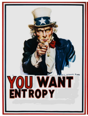


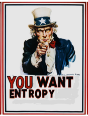


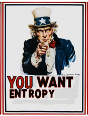


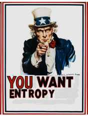
Take home message

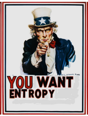
Entropy is useful for understanding projections.

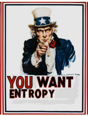


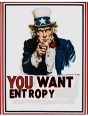


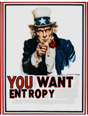


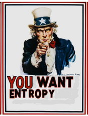


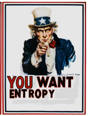












Proof

Examine the conditioned entropy

$$f(k) = H\big(d_{12} \,\big|\, d_{ij} \;\forall 1 \le i < j \le k, (i,j) \ne (1,2)\big).$$

Proof

Examine the conditioned entropy

$$f(k) = H\big(d_{12} \,\big|\, d_{ij}\ \forall 1 \le i < j \le k, (i,j) \ne (1,2)\big).$$

- $f(k)$ is decreasing (a general property of entropy).

Proof

Examine the conditioned entropy

$$f(k) = H\big(d_{12} \,\big|\, d_{ij}\ \forall 1 \le i < j \le k, (i,j) \ne (1,2)\big).$$

- $f(k)$ is decreasing (a general property of entropy).
- $f(1)$ and $f(\sqrt{n})$ are both ≈ 1 (not difficult).

Proof

Examine the conditioned entropy

$$f(k) = H\big(d_{12} \,\big|\, d_{ij} \; \forall 1 \le i < j \le k, (i,j) \ne (1,2)\big).$$

- $f(k)$ is decreasing (a general property of entropy).
- $f(1)$ and $f(\sqrt{n})$ are both ≈ 1 (not difficult).

Hence there exists some $k \le \sqrt{n}$ such that

$$f(k-1) - f(k) \le \frac{C}{\sqrt{n}}.$$

Proof

Examine the conditioned entropy

$$f(k) = H\big(d_{12}\,\big|\,d_{ij}\ \forall 1 \le i < j \le k, (i,j) \ne (1,2)\big).$$

- $f(k)$ is decreasing (a general property of entropy).
- $f(1)$ and $f(\sqrt{n})$ are both ≈ 1 (not difficult).

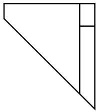
Hence there exists some $k \le \sqrt{n}$ such that

$$f(k-1) - f(k) \le \frac{C}{\sqrt{n}}.$$

Sandwich between $f(k-1)$ and $f(k)$ the following

$$f' := H\big(d_{12}\,\big|\,d_{ij}\ \{1 \le i < j \le k\} \setminus \{(1,2),(1,k),(2,k)\}\big)$$

and get $f' - f(k) \le C/\sqrt{n}$.



Proof II

$$H\Big(d_{12} \,|\, d_{i,j}\ \{1 \le i < j \le k\} \setminus \{(1,2),(1,k),(2,k)\}\Big) - H\Big(d_{12} \,|\, d_{i,j}\ \{1 \le i < j \le k\} \setminus \{(1,2)\}\Big) \le C/\sqrt{n}$$

Proof II

$$H\Big(d_{12}\,|\,d_{i,j}\ \{1 \le i < j \le k\} \setminus \{(1,2),(1,k),(2,k)\}\Big) - H\Big(d_{12}\,|\,d_{i,j}\ \{1 \le i < j \le k\} \setminus \{(1,2)\}\Big) \le C/\sqrt{n}$$

But "k" is just a label: we can label this vertex "3". So

$$H\Big(d_{12}\,|\,d_{i,j}\ \{1 \le i < j \le k\} \setminus \{(1,2),(1,3),(2,3)\}\Big) - H\Big(d_{12}\,|\,d_{i,j}\ \{1 \le i < j \le k\} \setminus \{(1,2)\}\Big) \le C/\sqrt{n}$$

Proof II

$$H\Big(d_{12}\,|\,d_{i,j}\ \{1 \le i < j \le k\} \setminus \{(1,2),(1,k),(2,k)\}\Big) - H\Big(d_{12}\,|\,d_{i,j}\ \{1 \le i < j \le k\} \setminus \{(1,2)\}\Big) \le C/\sqrt{n}$$

But "k" is just a label: we can label this vertex "3". So

$$H\Big(d_{12}\,|\,d_{i,j}\ \{1 \le i < j \le k\} \setminus \{(1,2),(1,3),(2,3)\}\Big) - H\Big(d_{12}\,|\,d_{i,j}\ \{1 \le i < j \le k\} \setminus \{(1,2)\}\Big) \le C/\sqrt{n}$$

This means that when you condition on "typical" values of d_{ij} for all $(i,j) \neq (1,2),(1,3),(2,3)$ you get

$$H(d_{12}) - H(d_{12}\,|\,d_{13},d_{23}) \le C/\sqrt{n}.$$

Csiszár's inequality

For any two variables X and Y, $H(X,Y) \leq H(X) + H(Y)$, with equality if and only if X and Y are independent.

Csiszár's inequality

For any two variables X and Y, $H(X, Y) \leq H(X) + H(Y)$, with equality if and only if X and Y are independent. Csiszár inequality is a quantitative version. It states that

$$\mathbb{E}_y\big(d_{TV}(X, \{X|Y = y\})^2\big) \leq H(X) - H(X|Y).$$

where d_{TV} is the total variation distance.

Csiszár's inequality

For any two variables X and Y, $H(X,Y) \leq H(X) + H(Y)$, with equality if and only if X and Y are independent. Csiszár inequality is a quantitative version. It states that

$$\mathbb{E}_y\big(d_{TV}(X, \{X|Y=y\})^2\big) \leq H(X) - H(X|Y).$$

where d_{TV} is the total variation distance. We now return to the metric polytope. Recall that we showed that, for typical values of d_{ij}, $(i,j) \neq (1,2)$, we have that

$$H(d_{12}) - H(d_{12} \,|\, d_{13}, d_{23}) \leq C/\sqrt{n}.$$

Csiszár's inequality

For any two variables X and Y, $H(X,Y) \le H(X) + H(Y)$, with equality if and only if X and Y are independent. Csiszár inequality is a quantitative version. It states that

$$\mathbb{E}_y\big(d_{TV}(X, \{X|Y=y\})^2\big) \le H(X) - H(X|Y).$$

where d_{TV} is the total variation distance. We now return to the metric polytope. Recall that we showed that, for typical values of d_{ij}, $(i,j) \neq (1,2)$, we have that

$$H(d_{12}) - H(d_{12} \,|\, d_{13}, d_{23}) \le C/\sqrt{n}.$$

By Csiszár's inequality, the total variation distance between d_{12} and $d_{12} \,|\, d_{13}, d_{23}$ is $\le Cn^{-1/4}$.

In other words: d_{12} is almost independent of d_{13}, d_{23}!

Csiszár's inequality

For any two variables X and Y, $H(X,Y) \leq H(X) + H(Y)$, with equality if and only if X and Y are independent. Csiszár inequality is a quantitative version. It states that

$$\mathbb{E}_y\big(d_{TV}(X, \{X|Y=y\})^2\big) \leq H(X) - H(X|Y).$$

where d_{TV} is the total variation distance. We now return to the metric polytope. Recall that we showed that, for typical values of d_{ij}, $(i,j) \neq (1,2)$, we have that

$$H(d_{12}) - H(d_{12} \,|\, d_{13}, d_{23}) \leq C/\sqrt{n}.$$

By Csiszár's inequality, the total variation distance between d_{12} and $d_{12} \,|\, d_{13}, d_{23}$ is $\leq Cn^{-1/4}$.

In other words: d_{12} is almost independent of d_{13}, d_{23}!

Since "1", "2" and "3" are just labels, we get that all three of d_{12}, d_{13} and d_{23} are almost independent (still conditioning on other d_{ij} lower than k).

Take home message

Entropy is useful for analyzing complicated dependency issues. Its monotonicty is crucial.

Back to metric spaces

Assume d_{12}, d_{13} and d_{23} are truly independent. Then, because *every* possible choice must satisfy the triangle inequality this implies conditions on the *supports*:

$$\operatorname{essmin} d_{12} + \operatorname{essmin} d_{23} \geq \operatorname{essmax} d_{13}$$

and its cousins.

Back to metric spaces

Assume d_{12}, d_{13} and d_{23} are truly independent. Then, because *every* possible choice must satisfy the triangle inequality this implies conditions on the *supports*:

$$\operatorname{essmin} d_{12} + \operatorname{essmin} d_{23} \geq \operatorname{essmax} d_{13}$$

and its cousins.

Hence

$$H(d_{12}, d_{13}, d_{23}) \leq \max \log(\underbrace{|I_{12}| \cdot |I_{13}| \cdot |I_{23}|}_{\text{the volume of the cube}})$$

where the maximum is taken over all triples I_{12}, I_{13}, $I_{23} \subset [0, 2]$ which satisfy the conditions on their minima and maxima.

Back to metric spaces

Assume d_{12}, d_{13} and d_{23} are truly independent. Then, because *every* possible choice must satisfy the triangle inequality this implies conditions on the *supports*:

$$\operatorname{essmin} d_{12} + \operatorname{essmin} d_{23} \geq \operatorname{essmax} d_{13}$$

and its cousins.

Hence

$$H(d_{12}, d_{13}, d_{23}) \leq \max \log(\underbrace{|I_{12}| \cdot |I_{13}| \cdot |I_{23}|}_{\text{the volume of the cube}})$$

where the maximum is taken over all triples I_{12}, I_{13}, $I_{23} \subset [0, 2]$ which satisfy the conditions on their minima and maxima. A little calculation (which we will not do) shows that the maximum is achieved when $I_{12} = I_{13} = I_{23} = [1, 2]$.

Metric spaces II

Since d_{12}, d_{13} and d_{23} are not truly independent but only conditionally almost independent, we get

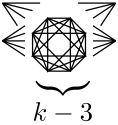
$$H(d_{12}, d_{13}, d_{23} \,|\, d_{i,j}\ \forall 1 \leq i < j \leq k, 4 \leq j) \leq Cn^{-c}.$$

Metric spaces II

Since d_{12}, d_{13} and d_{23} are not truly independent but only conditionally almost independent, we get

$$H(d_{12}, d_{13}, d_{23} \,|\, d_{i,j}\ \forall 1 \leq i < j \leq k, 4 \leq j) \leq Cn^{-c}.$$

Let $A = \{(i,j) : i < j, i \leq k-3\}$, which we think about as the core (all edges between $\{1, \ldots, k-3\}$) and spikes.



Metric spaces II

Since d_{12}, d_{13} and d_{23} are not truly independent but only conditionally almost independent, we get

$$H(d_{12}, d_{13}, d_{23} \,|\, d_{i,j} \; \forall 1 \leq i < j \leq k, 4 \leq j) \leq Cn^{-c}.$$

Let $A = \{(i,j) : i < j, i \leq k-3\}$, which we think about as the core (all edges between $\{1, \ldots, k-3\}$) and spikes. For any $a, b, c > k$ we have

$$H(d_{ab}, d_{bc}, d_{ac} \,|\, A) \leq$$
$$\leq H(d_{ab}, d_{bc}, d_{ac} \,|\, d_{ij} \; \forall i \leq k-3, j \in \{1, \ldots, k-3, a, b, c\})$$

because of conditional entropy monotonicity.

Metric spaces II

Since d_{12}, d_{13} and d_{23} are not truly independent but only conditionally almost independent, we get

$$H(d_{12}, d_{13}, d_{23} \,|\, d_{i,j} \;\forall 1 \leq i < j \leq k, 4 \leq j) \leq Cn^{-c}.$$

Let $A = \{(i,j) : i < j, i \leq k-3\}$, which we think about as the core (all edges between $\{1, \ldots, k-3\}$) and spikes. For any $a, b, c > k$ we have

$$\begin{aligned} &H(d_{ab}, d_{bc}, d_{ac} \,|\, A) \leq \\ &\quad \leq H(d_{ab}, d_{bc}, d_{ac} \,|\, d_{ij} \;\forall i \leq k-3, j \in \{1, \ldots, k-3, a, b, c\}) \\ &\quad \leq Cn^{-c} \end{aligned}$$

because of conditional entropy monotonicity. Relabeling a, b and c to 1, 2 and 3, and $\{1, \ldots, k-3\}$ to $\{4, \ldots, k\}$ we are back in what we know.

Metric spaces III

$A = \{(i,j) : i < j, i \le k-3\}$ $\qquad$ $H(d_{ab}, d_{bc}, d_{ac} \,|\, A) \le Cn^{-c}$

We use this for all a, b and $c > k-3$ and Shearer's inequality[1] to get

$$H(\{d_{ab}\}_{a,b \ge k-3} \,|\, A) \le Cn^{2-c}.$$

[1] Alternatively one can choose a Steiner system of triangles and avoid Shearer's inequality

Metric spaces III

$A = \{(i,j) : i < j, i \leq k-3\}$ $\qquad$ $H(d_{ab}, d_{bc}, d_{ac} \,|\, A) \leq Cn^{-c}$

We use this for all a, b and $c > k-3$ and Shearer's inequality[1] to get

$$H(\{d_{ab}\}_{a,b \geq k-3} \,|\, A) \leq Cn^{2-c}.$$

Adding the entropy of the distances in A is not a problem as they are bounded by $C|A| < Cnk \leq Cn^{3/2}$. So the total entropy of the metric polytope is $\leq Cn^{2-c}$. □

[1] Alternatively one can choose a Steiner system of triangles and avoid Shearer's inequality

The proof in a nutshell

- Use the monotonicity of conditional entropy and Csiszár's inequality to show that it is enough to condition on a small ($\leq \sqrt{n}$) number of vertices to get that three fixed distances are almost independent.
- Find the optimal solution under the combined conditions of metricity and independence.
- Condition on $n^{3/2}$ edges to get almost independence for all triples, and use Shearer's inequality for the conditioned measure. Get that the total entropy is $\leq n^{3/2} + n^{2-c}$.

Miscellanea

- The vertices of the metric polytope are *exchangeable*. This gave us the first big tip that one should expect conditional independence.

Miscellanea

- The vertices of the metric polytope are *exchangeable*. This gave us the first big tip that one should expect conditional independence. Using *weak exchangeability* (a.k.a. joint exchangeablity for arrays) one can prove that $\operatorname{vol} M_n = e^{o(n^2)}$ but the o is completely unexplicit.

Miscellanea

- The vertices of the metric polytope are *exchangeable*. This gave us the first big tip that one should expect conditional independence. Using *weak exchangeability* (a.k.a. joint exchangeablity for arrays) one can prove that $\operatorname{vol} M_n = e^{o(n^2)}$ but the o is completely unexplicit.
- The Szemerédi regularity lemma gives another way to prove conditional independence.

Miscellanea

- The vertices of the metric polytope are *exchangeable*. This gave us the first big tip that one should expect conditional independence. Using *weak exchangeability* (a.k.a. joint exchangeablity for arrays) one can prove that $\operatorname{vol} M_n = e^{o(n^2)}$ but the o is completely unexplicit.
- The Szemerédi regularity lemma gives another way to prove conditional independence. Using this approach one can give an explicit o, but it involves inverse super-tower functions.

Miscellanea

- The vertices of the metric polytope are *exchangeable*. This gave us the first big tip that one should expect conditional independence. Using *weak exchangeability* (a.k.a. joint exchangeablity for arrays) one can prove that $\operatorname{vol} M_n = e^{o(n^2)}$ but the o is completely unexplicit.
- The Szemerédi regularity lemma gives another way to prove conditional independence. Using this approach one can give an explicit o, but it involves inverse super-tower functions.
- We also have a proof based on the Kővári-Sós-Turán theorem that gives $\operatorname{vol} M_n \leq C \exp(Cn^2/\log^c n)$.

Thank you