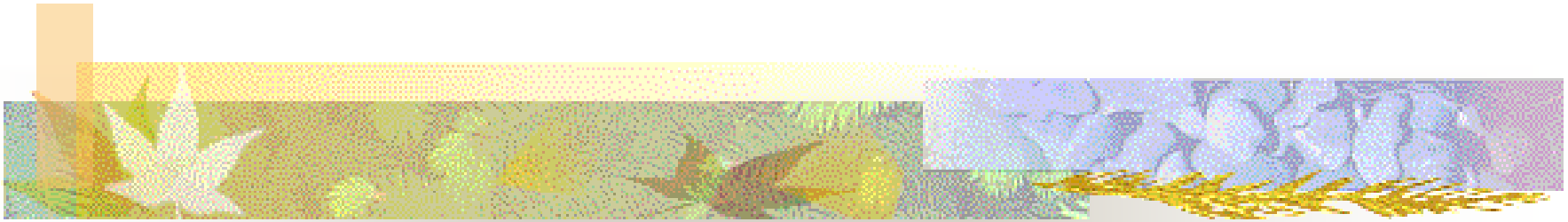
大規模知識処理特論 (第1回)
ガイダンス



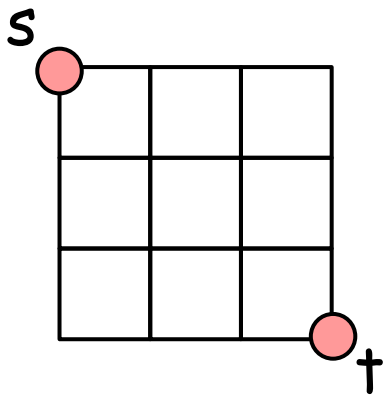
北海道大学 情報科学研究院
堀山 貴史
脊戸 和寿

大規模知識処理特論

- この授業の目標
  - 知識の編集・分類・解析・索引化等の知的な情報処理を行うために不可欠な知識処理の技法について学びます
- 大規模知識処理について、以下の観点から
  並行して or 順に 学習を進めます
  - 最適化技法
  - 論理関数と計算量理論の基礎
  - 厳密アルゴリズムと近似アルゴリズム
  - BDD/ZDDによる離散構造処理
- 適当な時期にレポート課題を課します

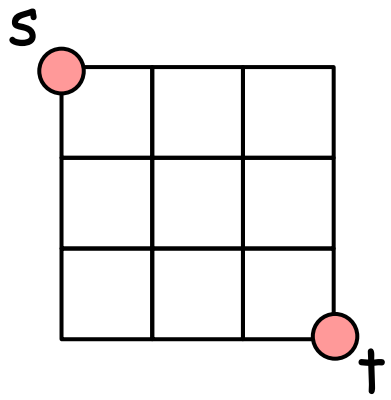
Quiz： s-t パス (最短路)

Q： s から t への**最短路**を列挙しなさい (何通り？)



Quiz： s-t パス (最短路)

Q： s から t への**最短路**を列挙しなさい (何通り？)



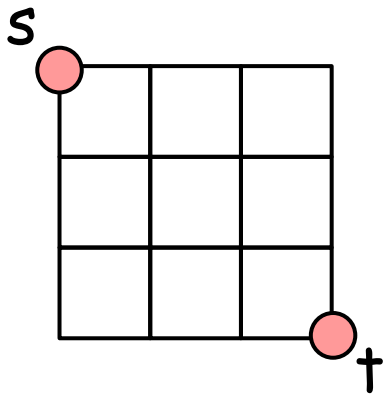
の組合せで、1 つの最短路

最短路は、$_6C_3 = \dfrac{6 \cdot 5 \cdot 4}{3 \cdot 2 \cdot 1} = 20$ 通り

Quiz： s-t パス ~~(最短路)~~

Q： s から t への**経路**を列挙しなさい (何通り？)

- **遠回り ok**
- **同じ点**を通るのは **NG**



- 何通りかを求める公式は、ありません

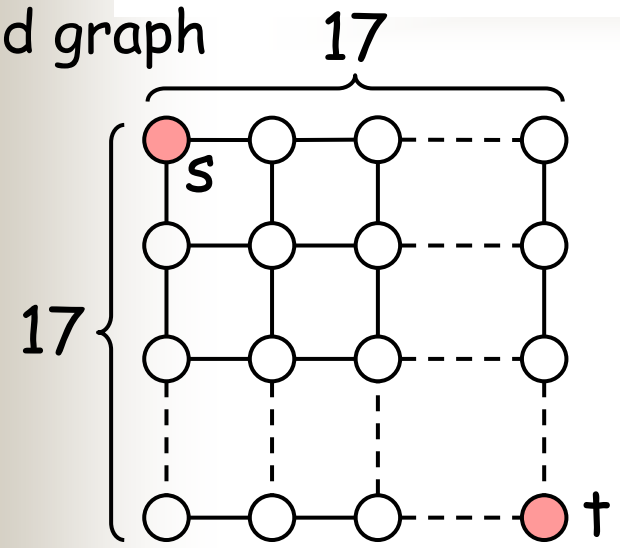
(ちなみに、答えは 184 通り)

列挙： 条件を満たすものを全部

- 求められること
  - **効率的に**列挙したい (時間計算量)
  - 結果を**コンパクト**に持ちたい (領域計算量)
  - 結果を**簡単に**利用したい
    - 何個ある？ / サンプリングしたい
    - 後で条件を色々変えて、結果を検索したい

例）s-t パス（同じ点を2回以上 通らない）

Grid graph



s-t パスの個数　6,344,814,
611,237,963,971,310,297,540,
795,524,400,449,443,986,866,
480,693,646,369,387,855,336

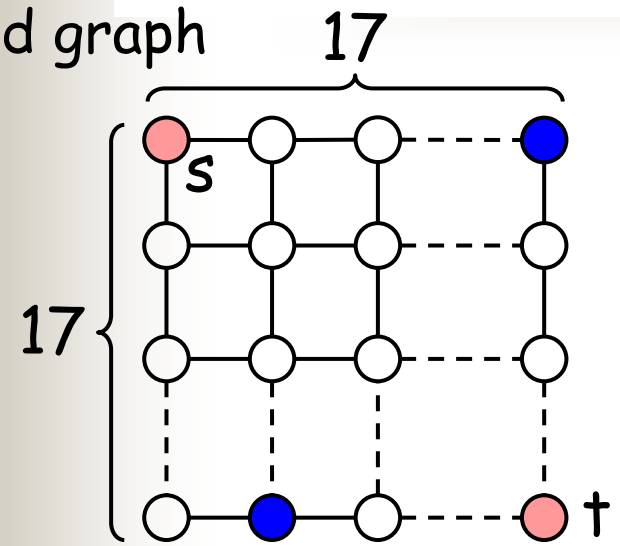
約 6.3×10^{61}　(63 那由他)

サンスクリット語で
“十分に大きな数”

- 求められること
  - **効率的に**列挙したい (時間計算量)
  - 結果を**コンパクト**に持ちたい (領域計算量)
  - 結果を**簡単に**利用したい
    - 何個ある？ / サンプリングしたい
    - 後で条件を色々変えて、結果を検索したい

例）s-t パス（同じ点を2回以上 通らない）

Grid graph



s-t パスの個数　6,344,814,
611,237,963,971,310,297,540,
795,524,400,449,443,986,866,
480,693,646,369,387,855,336

約 6.3×10^{61}　(63 那由他)

サンスクリット語で
“十分に大きな数”

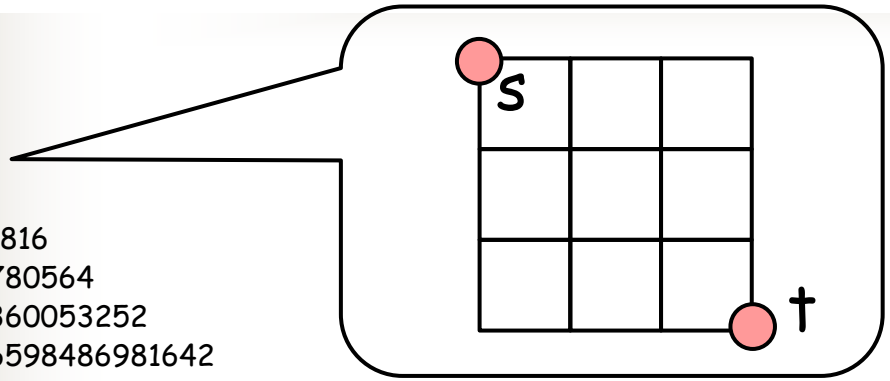
- 求められること
  - **効率的に**列挙したい (時間計算量)
  - 結果を**コンパクト**に持ちたい (領域計
  - 結果を**簡単に**利用したい
    - 何個ある？ / サンプリン
    - 後で条件を色々変えて、結果を検索したい

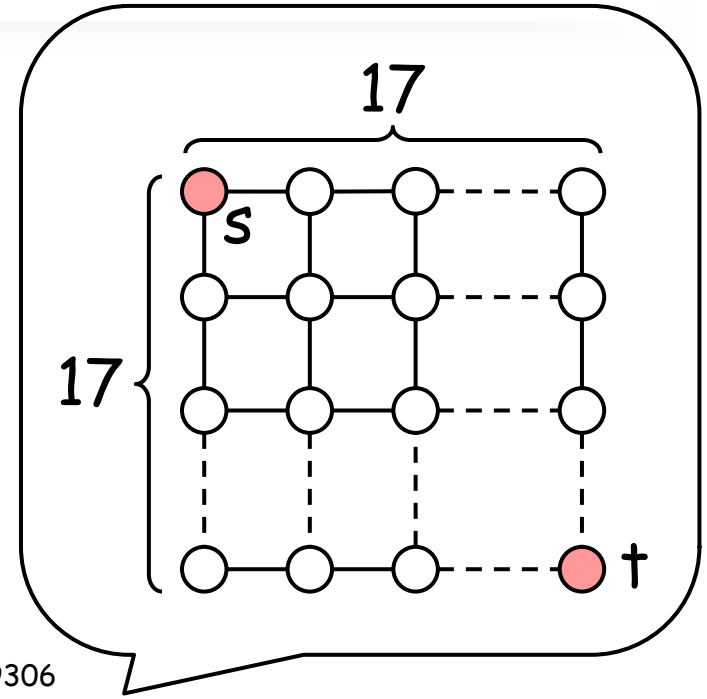
長さ50以下の
パスは何個？

最長の
パスは？

● を2つとも通る
パスは何個？

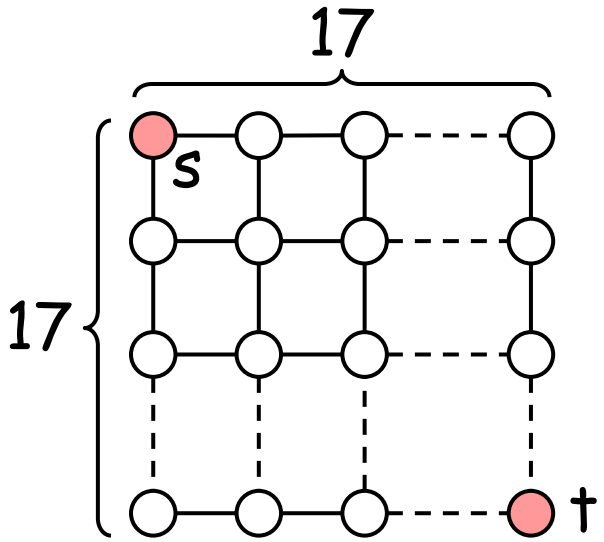
s-t パスは、何通り？





- 1 2
- 2 12
- 3 184
- 4 8512
- 5 1262816
- 6 575780564
- 7 789360053252
- 8 3266598486981642
- 9 41044208702632496804
- 10 1568758030464750013214100
- 11 182413291514248049241470885236
- 12 64528039343270018963357185158482118
- 13 69450664761521361664274701548907358996488
- 14 22744971467681273963182645932798963387613323440
- 15 22667455688626727463745673967130989348663248854083190 28
- 16 68745445609149931587631563132489232824587945968099457285419306
- 17 63448146112379639713102975407955244004494439868664806936463693878553 36
- 18 17821128408420651298933849466523252751678380657047676559314524746058266927825 32
- 19 15233449717048799930807428103192296908994542553232945557760298667373550605928775692558 44
- 20 396289219982303756020729951713336250210633970573946377151523711337701068236403570670447206494039 8

s-t パスは、何通り？



- 21
  313747510501371027204205381373822145131033121936987236530613519913464333793893857939655769922460213164638 68
- 22
  7559702866673453396615191233152226193531037320724094811673914104795179257927436312349870388833176349872 71171404439792
- 23
  5543542935523747700991431848906143793069037997096433133255695864648400840733488554456638692402087571124206008540851348293394572 0
- 24
  1237171223120706475833874486267357083237304198901294353967872708048495169551593048564139455079215303719185802821251228092660030458 1386791094
- 25
  8402974857881133471007083745436809127296054293775383549824742623937028497898215256929178577083970960121625602506027316549718402106494049978375604247408
- 26
  1736993158627927293117544042123649890037222958828814060466370372091034241327613476278921819349800610708229622314338049134829002672193112962770873889085390810890639 6

大規模知識処理アルゴリズム

- **問題の解き方** (ざっくり言うと 作戦) を考える
- どうして アルゴリズムは **重要** ？
  - **実装**を頑張って、スピードが **2倍，3倍** …
    アルゴリズムを変えると **1000倍** …

    → より大規模/実用的な対象が扱える
  - **基礎理論**が、様々な**応用分野**とつながる

    例)
    - 「選挙区の区割り」と「避難所の割当て」
    - 「展開図」と「細胞で折り紙」

    など

大阪府だと、489穣3281秄3930垓3925京22兆5190億1210万1206通りの区割り
（普通にやると、ハードディスクに入らない）
0.34秒で求める

選挙区の区割り

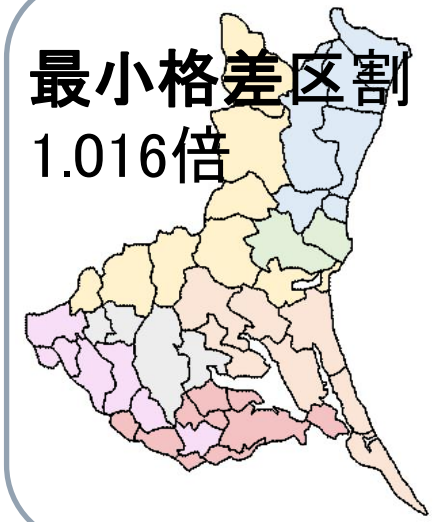
- **一票の格差**

例: 茨城県

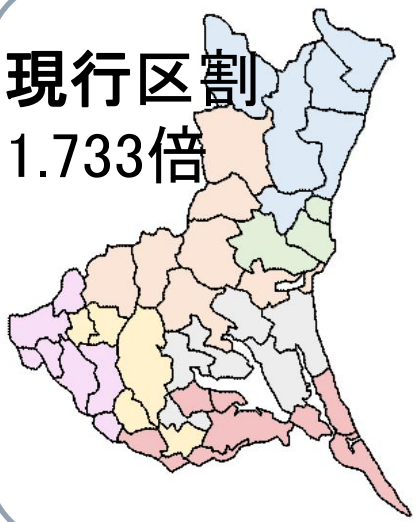
41市区郡
7選挙区

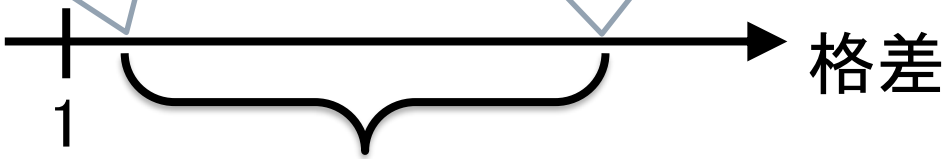
国勢調査の結果をもとに、
格差が小さくなるように、区割りをする

最小格差区割
1.016倍



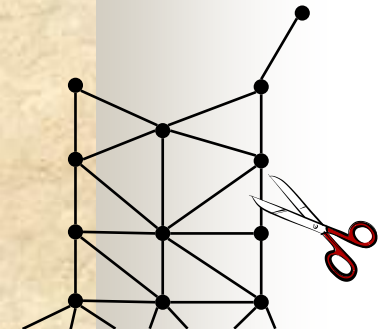
現行区割
1.733倍





この間の様子はどうなっているのか
解は多いのか、少ないのか

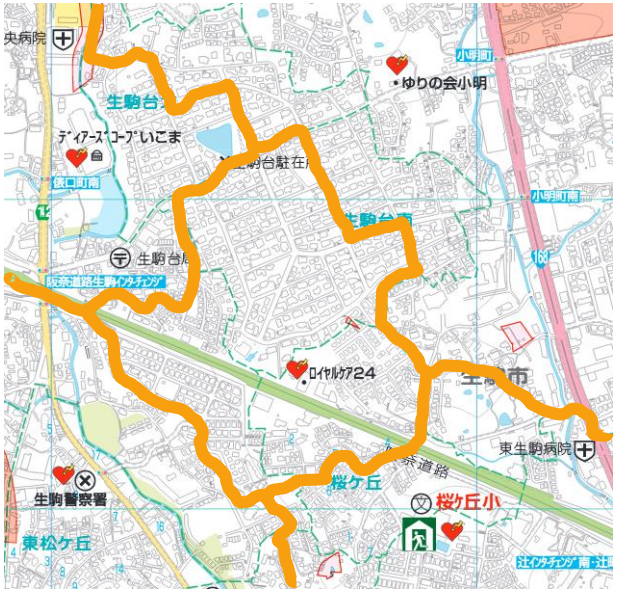
問題の取り組み方
- 一気に考えるのは大変
- 少しずつ小さな目標に分ける



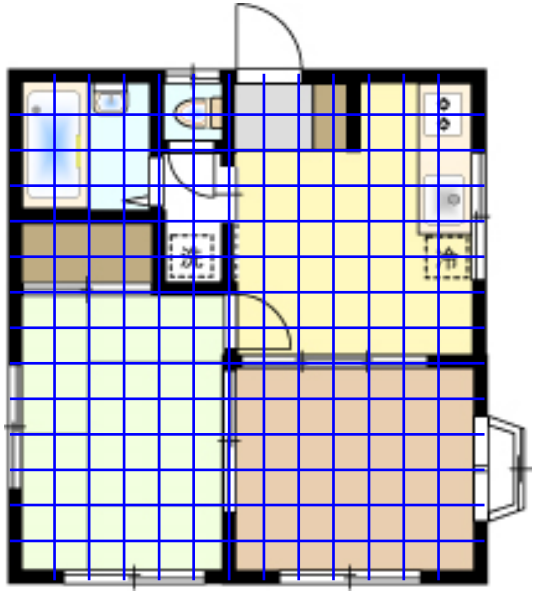
たとえば、
グラフを切り分ける

選挙区の区割りの応用

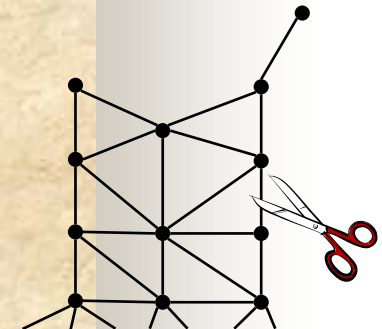
- 同様のアイデアが、**他の分野にも応用**できる



被災時の避難計画
（避難場所の割当）



フロアプラン
（間取り図）



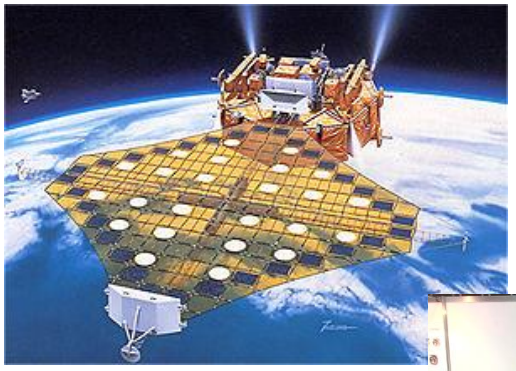
たとえば、
グラフを切り分ける

Folding Mechanism (折りメカニズム)

商品のパッケージ

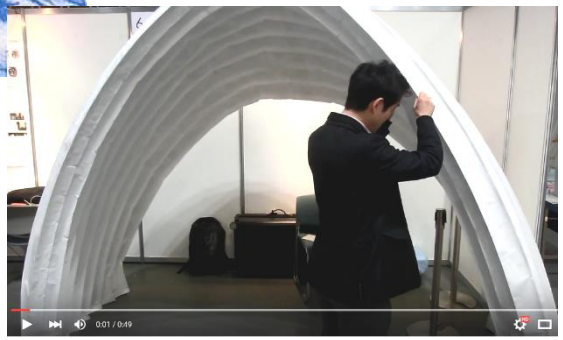
Folding Mechanism (折りメカニズム)

商品のパッケージ

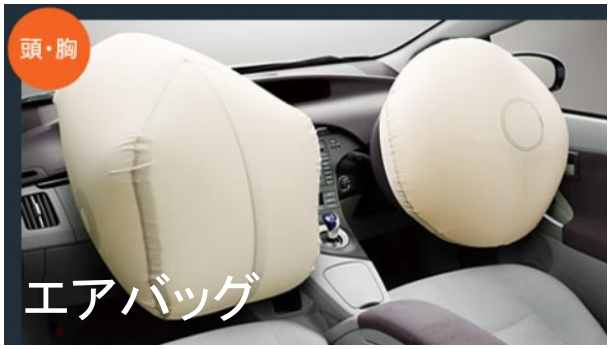


人工衛星のパネル
(ミウラ折り)

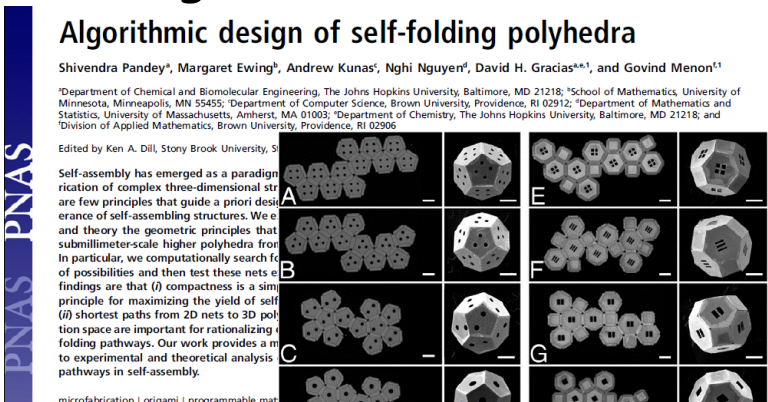
建築

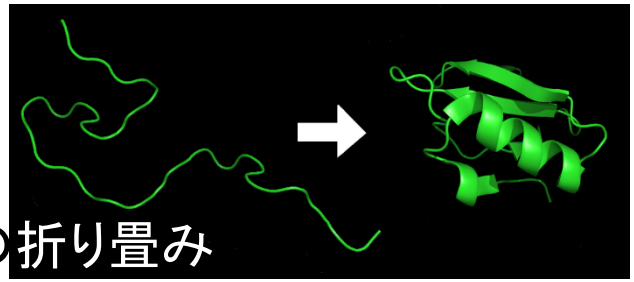


缶飲料
(ヨシムラ パターン)



Origami 研究に、大型予算も

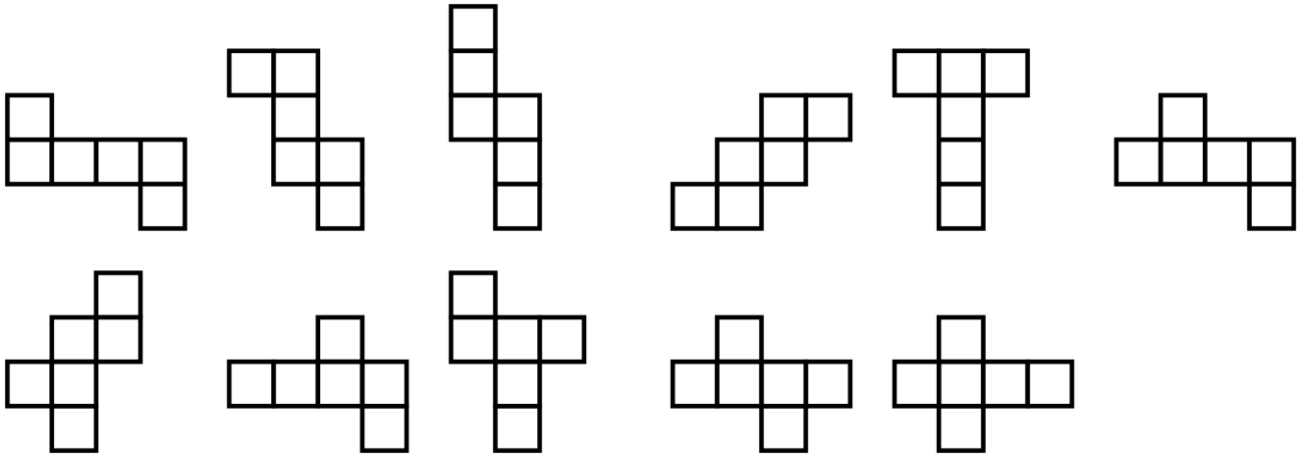




タンパク質の折り畳み

展開図の研究

サイコロの展開図は 11 種類

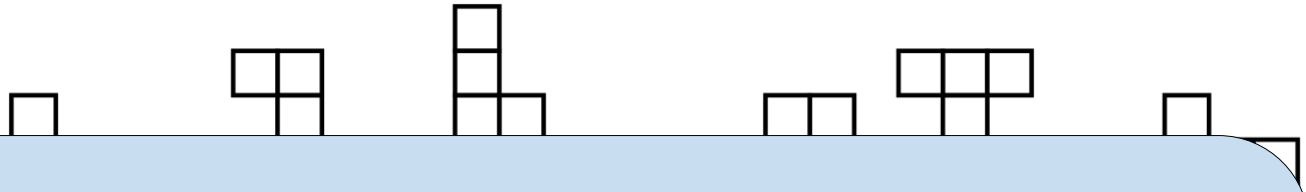


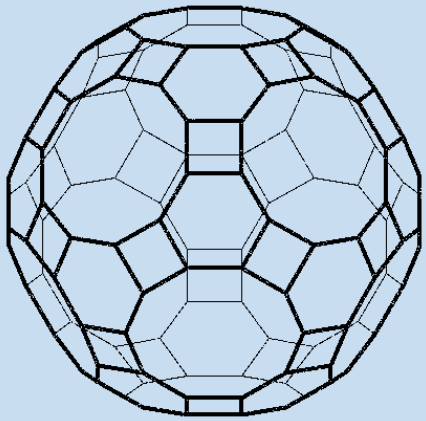
Q.2 サッカーボールの展開図は何種類？

A. 3,127,432,220,939,473,920
（312京 7432兆 2209億 3947万 3920）

展開図の研究

サイコロの展開図は 11 種類





181,577,189,197,376,045,
928,994,520,239,942,164,480

181澗 5771溝 8919穣 7376秭 459垓
2899京 4520兆 2399億 4216万 4480

A. 3,127,432,220,939,473,920
（312京 7432兆 2209億 3947万 3920）

この授業では：　離散構造の観点から

- **アルゴリズム**で、社会を支える
  （列挙/最適化 アルゴリズムを中心に）
- **社会とのつながり**は色々

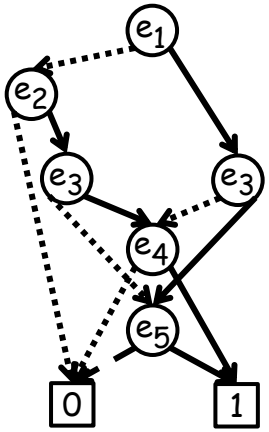
離 散 構 造

離散構造（組合せ、グラフなど）を
どう表し、どう扱うか

離散構造の表現

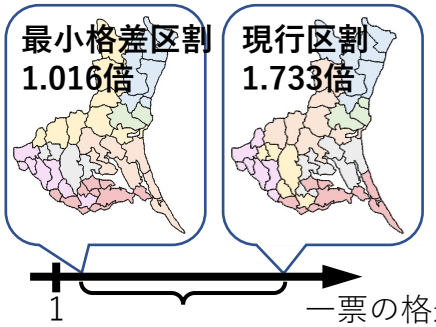
構築アルゴリズム

$$\{\ \{e_1\ e_3\ e_5\},\ \{e_2\ e_3\ e_4\},\ \{e_1\ e_4\},\ \{e_2\ e_5\}\ \}$$



分野横断的なつながり

選挙区の区割り



被災時の避難計画

避難場所の割当て

計算折り紙

細胞の形成への応用