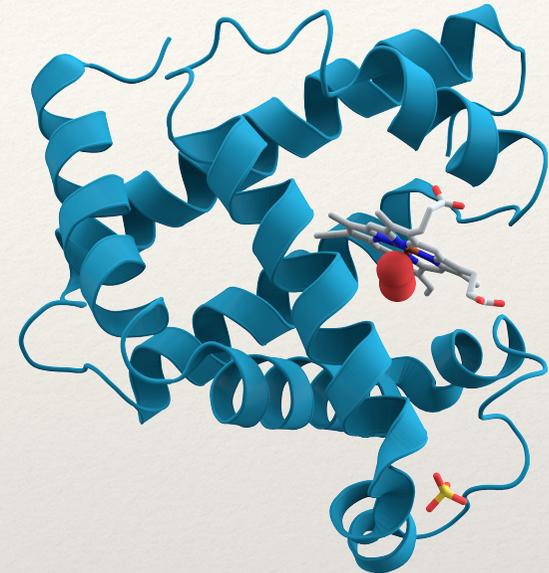


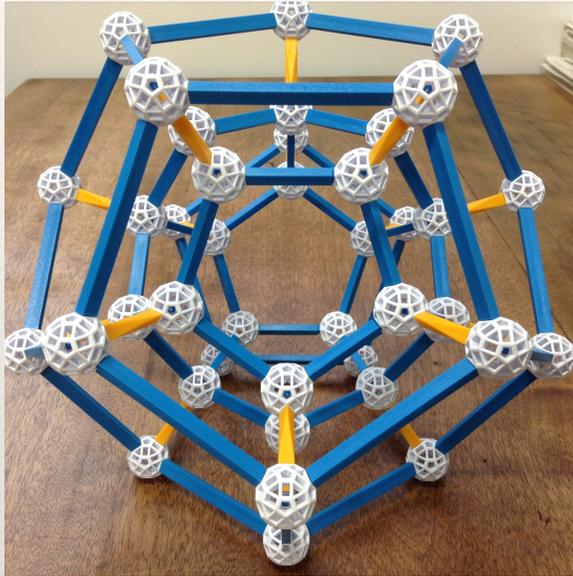
# 幾つかの応用の話題

## トポロジーを用いたデータ解析手法

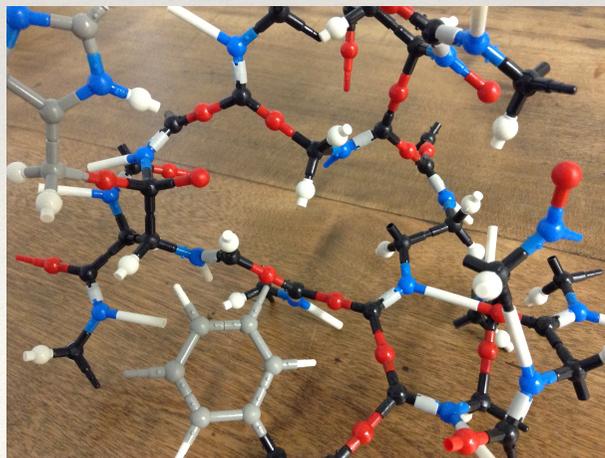
- ❖ タンパク質の幾何解析
- ❖ ガラスの幾何解析
- ❖ センサーの幾何解析



# 原子配置と幾何



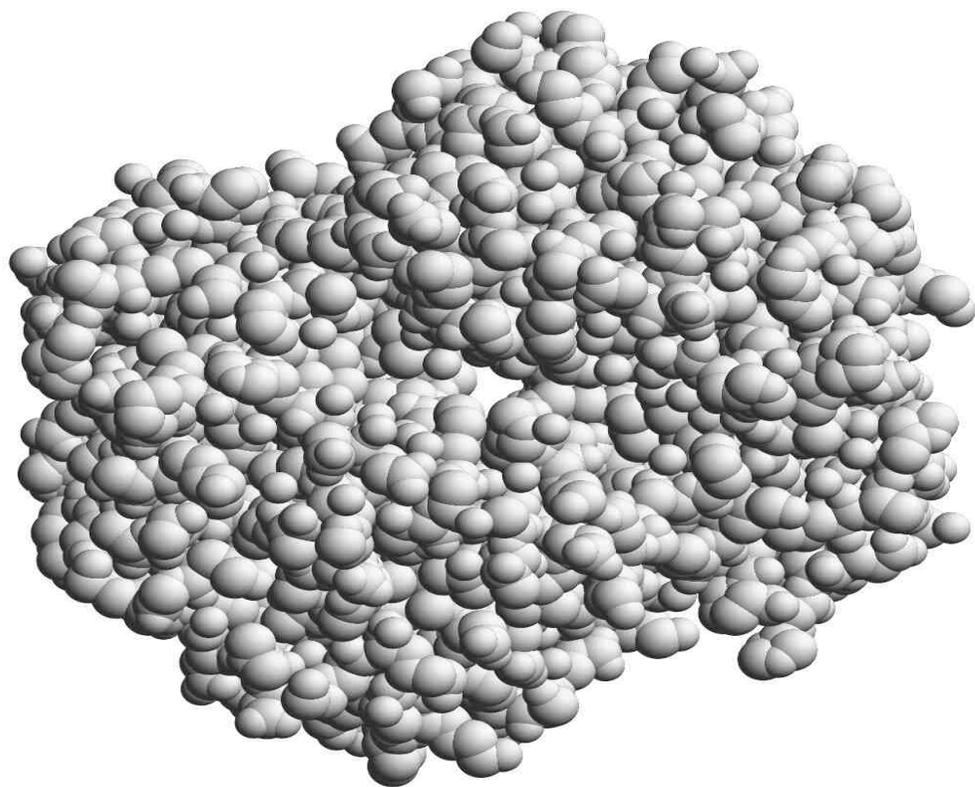
- ❖ 結晶構造：群論，組合せ論，  
離散幾何 etc.



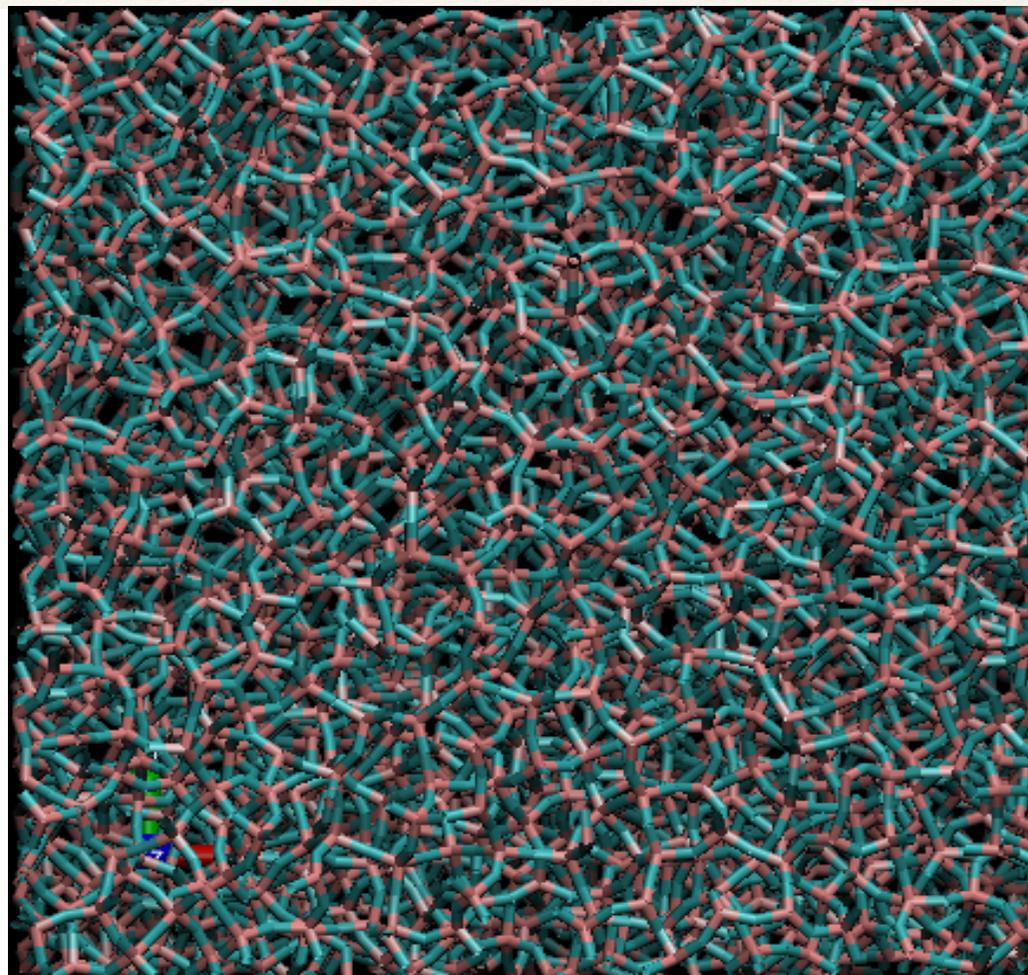
- ❖ 不規則構造 (タンパク質, ガラス)：  
適切な特徴付け (言語) ???

# 原子配置と幾何

ヘモグロビン

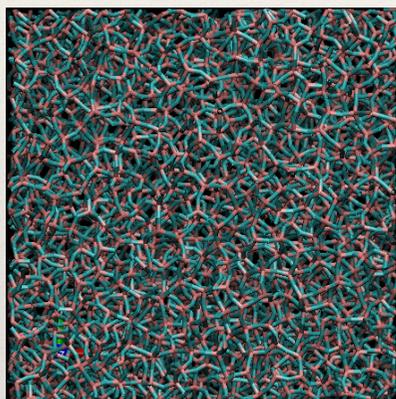


ガラス



# データから情報を取り出す

幾何データ



入力



出力

???

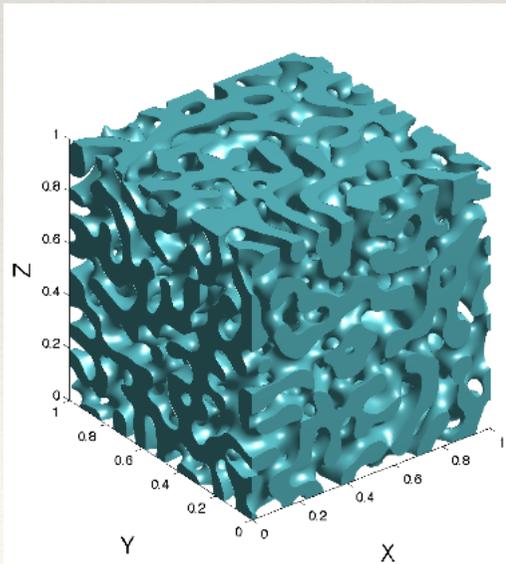
## 要求

- ❖ 形の特徴を抽出して欲しい
- ❖ 理解しやすいものが良い
- ❖ 計算（実用化）可能なものが欲しい
- ❖ 膨大なデータに対して処理したい

# ホモロジー



入力



出力

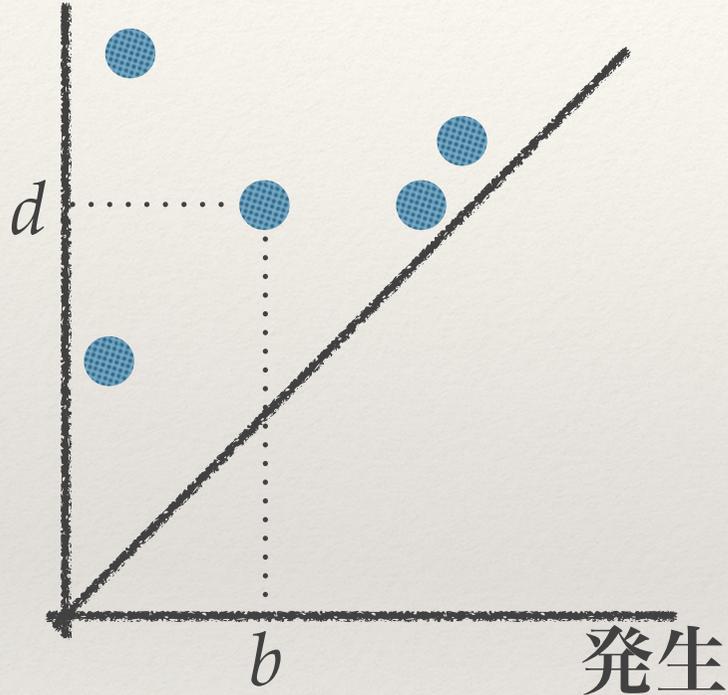
$$H_0(\text{cube}) = \mathbf{R} \quad : 1 \text{ 個の連結成分}$$

$$H_1(\text{cube}) = \mathbf{R}^{847} \quad : 847 \text{ 個のリング}$$

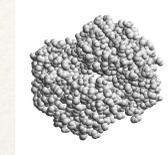
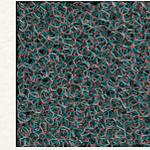
$$H_2(\text{cube}) = 0 \quad : 0 \text{ 個の空洞}$$

# パーシステントホモロジー

消滅



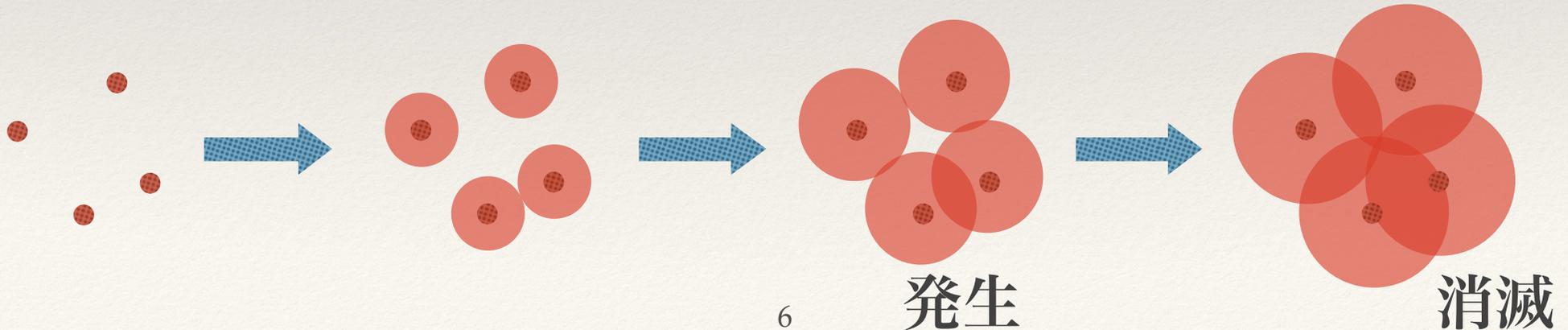
❖ 入力: 幾何データ



❖ 出力: 平面内の点の集まり

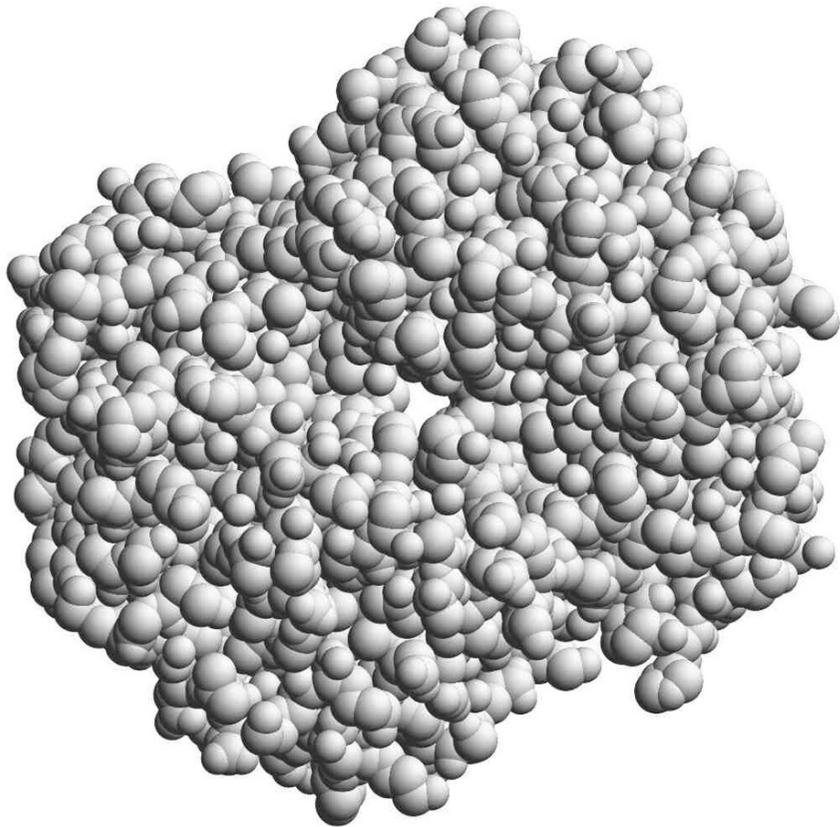
❖ 各点●は  $l$  次元の穴に対応

❖  $b$  &  $d$  は発生・消滅の半径スケール



# タンパク質への応用

ヘモグロビン



(約 5000 個の原子)

立体構造 ↔ 機能

立体構造データベース

RCSB **PDB**  
PROTEIN DATA BANK

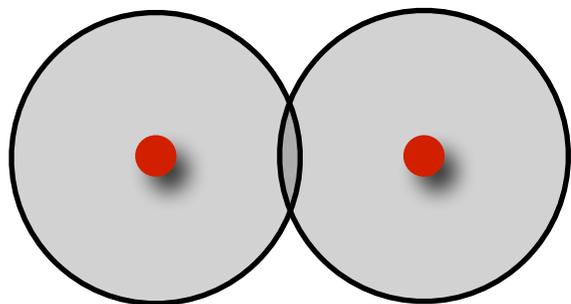
(約 100,000 個のデータ)

目標

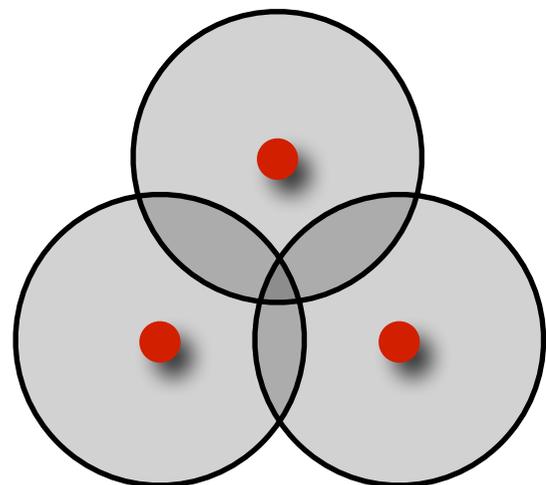
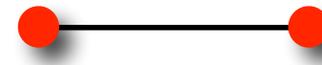
トポロジーを用いた  
データ解析手法の開発

# タンパク質への応用

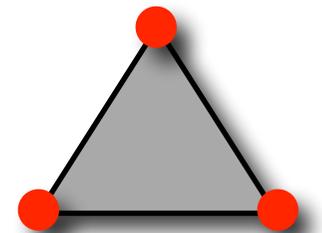
- 各原子を適当な球で表現



2つの原子球が  
共通部分をもて  
ば辺を張る



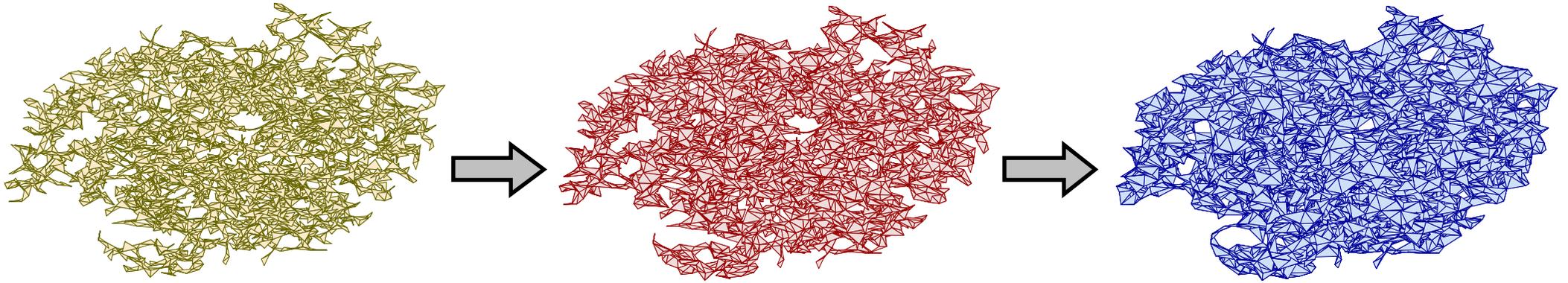
3つの原子球が  
共通部分をもて  
ば三角形を張る



以下同様にして作られる多面体を Čech複体 とよぶ

**脈体定理**：Čech複体はタンパク質球体モデルの  
トポロジーに関する性質を保つ

# タンパク質への応用



ヘモグロビンの Čech 複体増大列

タンパク質内に現れるロバストな穴を取り出したい

➡ パーシステントホモロジー群

# タンパク質の幾何解析

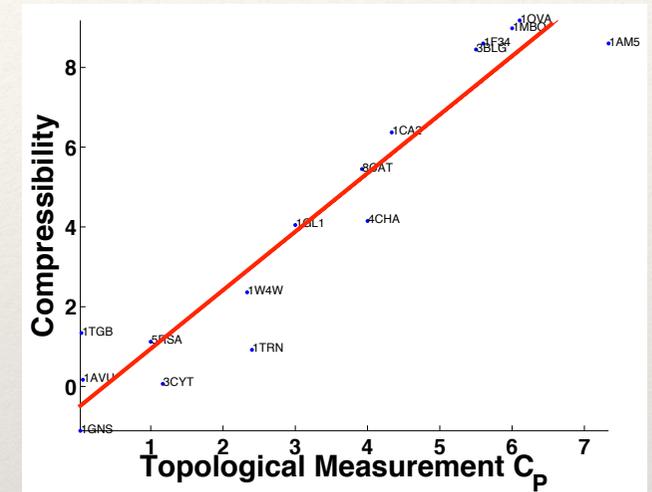
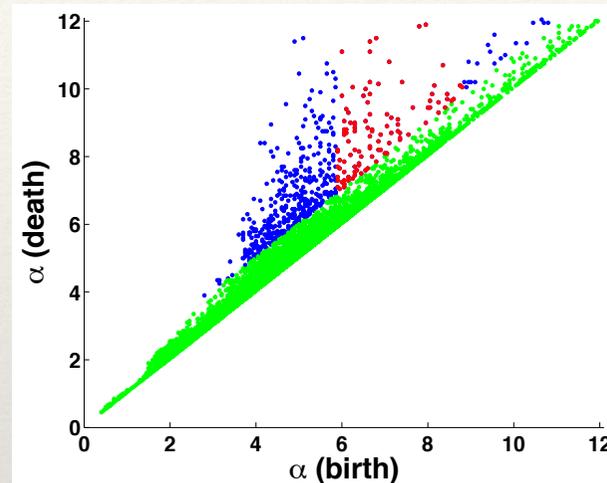
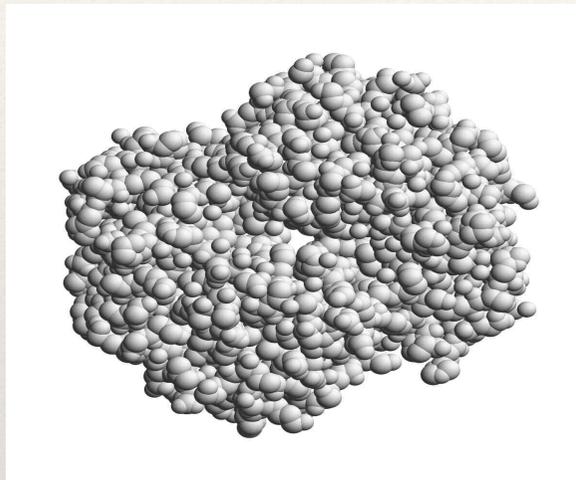
タンパク質



ホモロジー計算

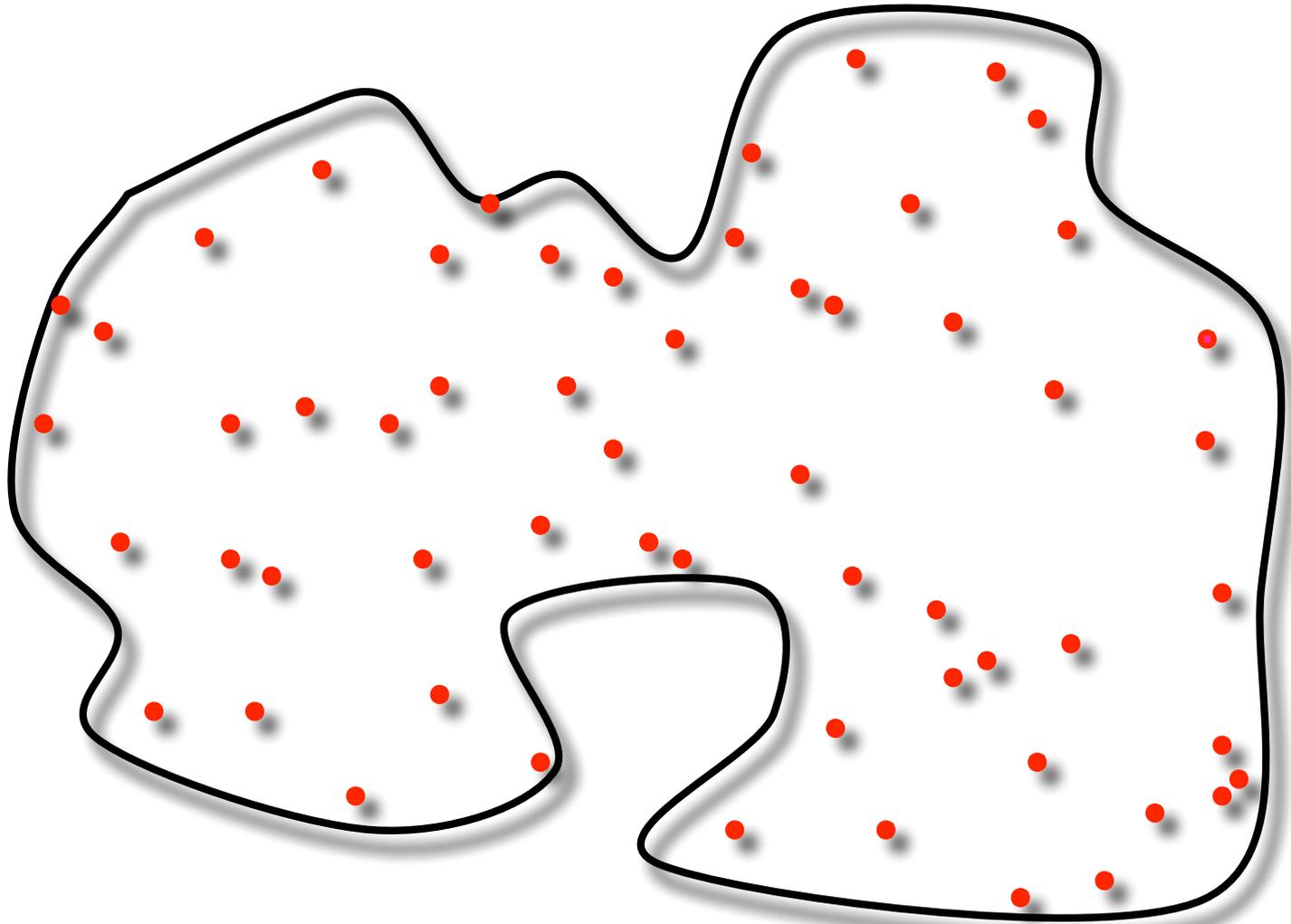


圧縮率推定

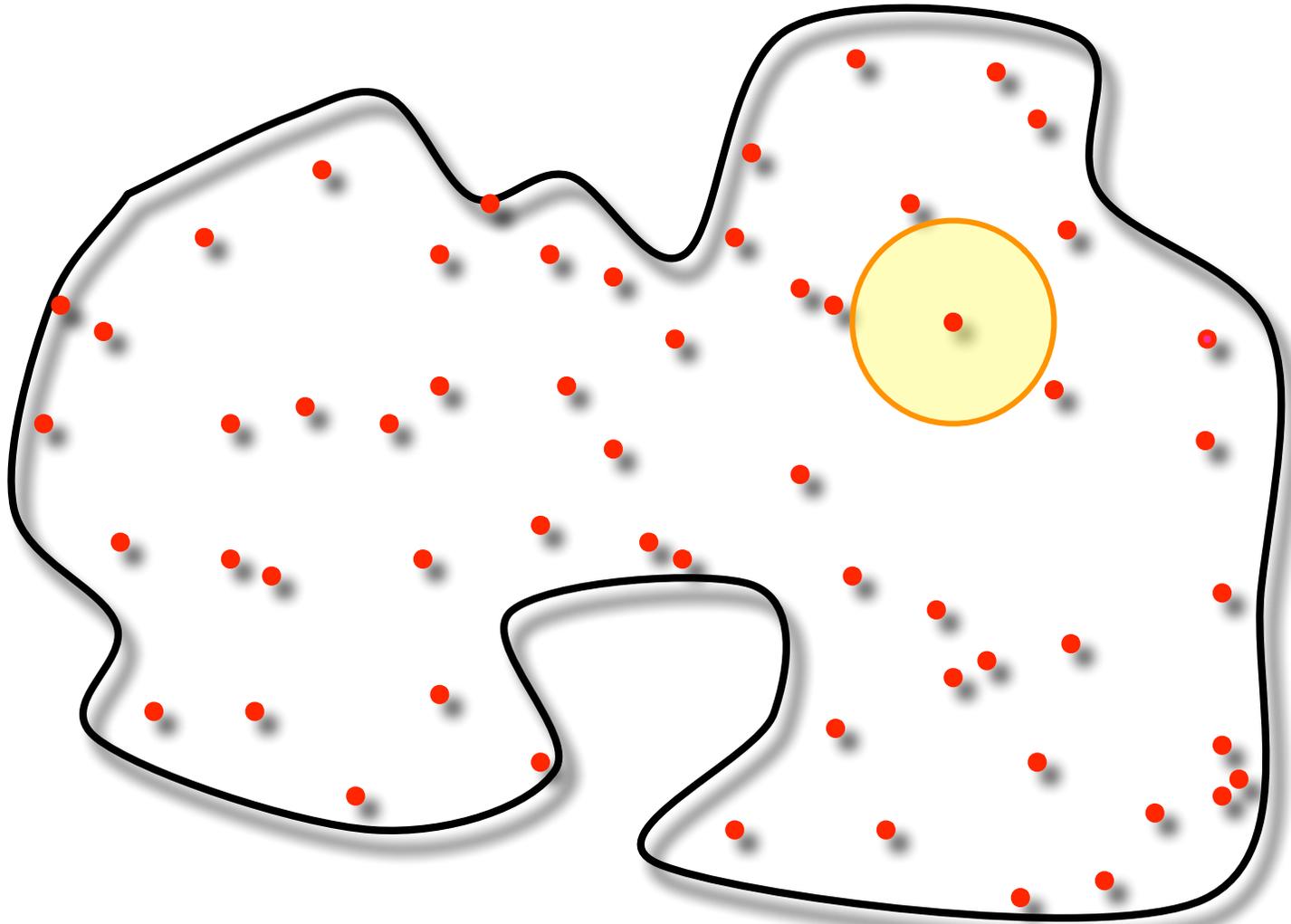


- ❖ **穴の数**だけでなく、**形**についての情報も手に入る  
(RosettaHoles等より情報豊富)
- ❖ 幾何構造に基づく分類 (系統樹)
- ❖ フォールディング問題

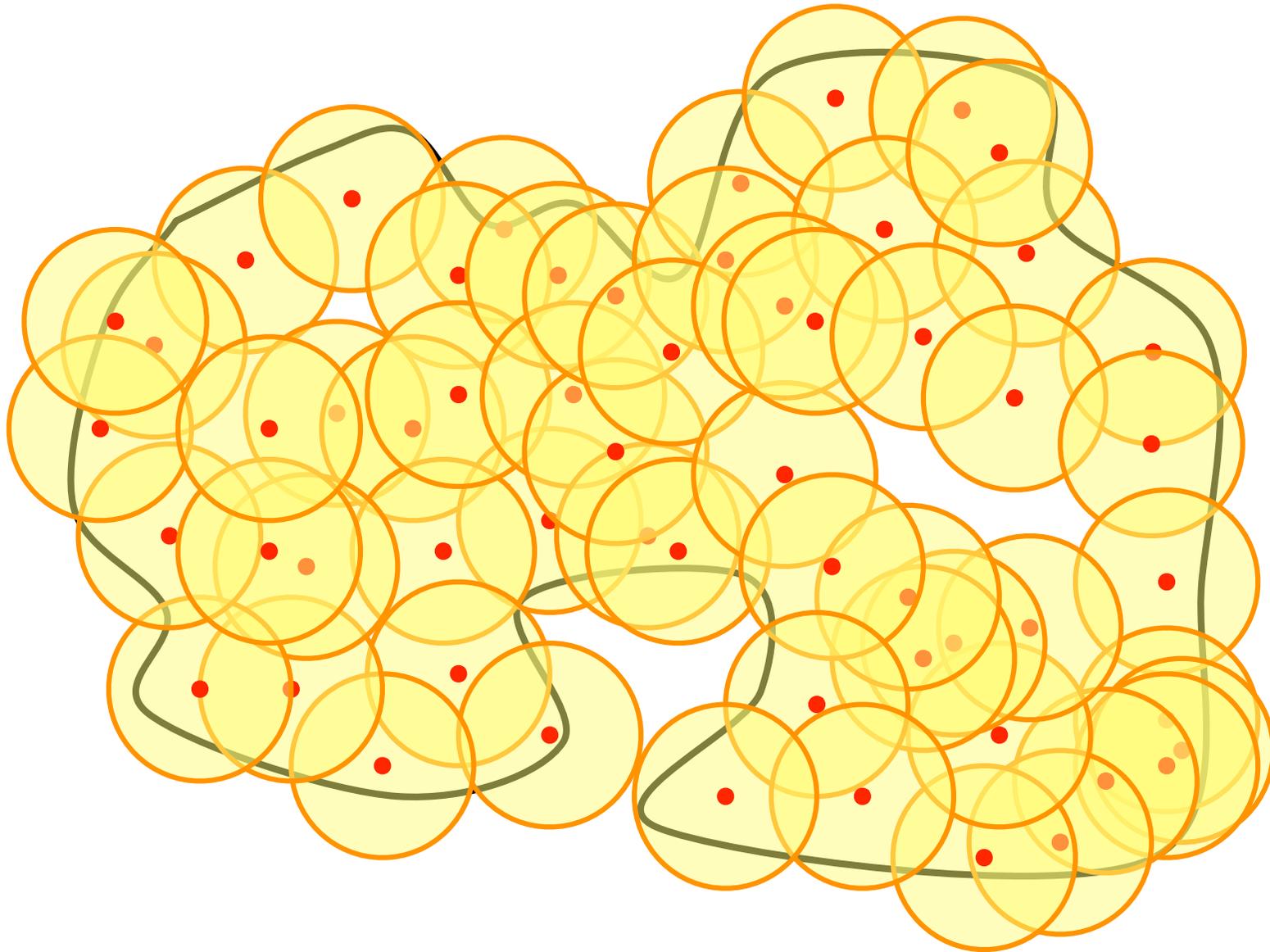
# センサーネットワーク



# センサーネットワーク



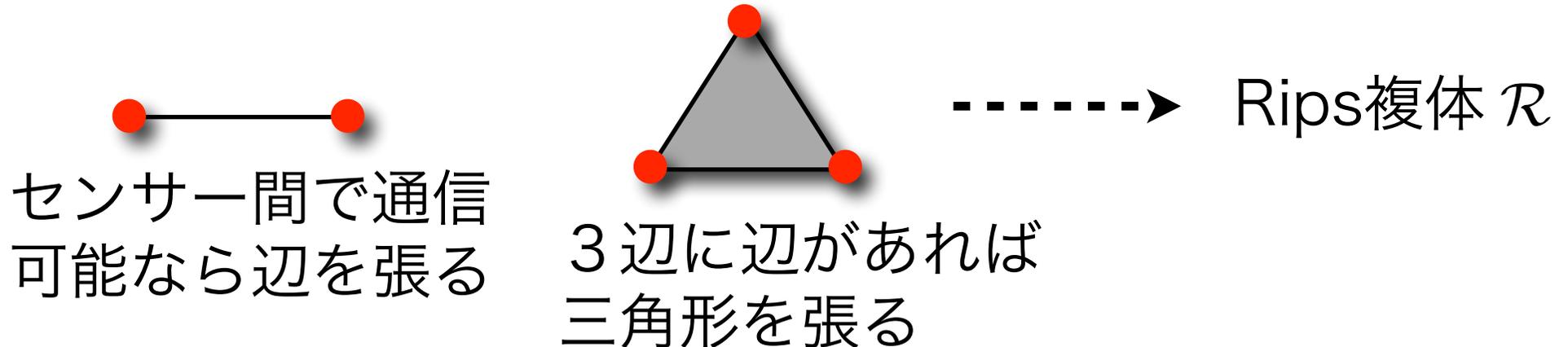
# センサーネットワーク



被覆問題：センシング領域は対象領域を覆っているか？

# センサーネットワーク

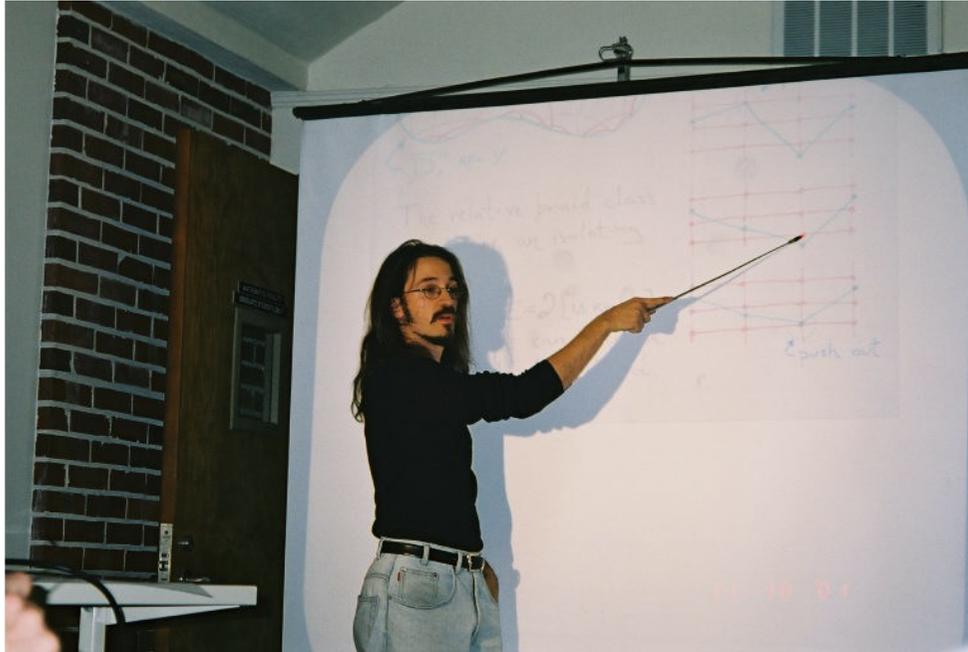
- センサーは低性能（センシング半径  $r_s$ ，通信半径  $r_c$  は小さい．特に  $r_s \geq r_c/\sqrt{3}$ ）
- センサーの絶対位置はわからない  
→  $r_s$  を直接使えない！
- センサー間の通信情報から多面体（Rips複体）を構成（局所情報から構成可能）



# センサーネットワーク

- 被覆定理 (de Silva & Ghrist)

$\delta[\sigma] \neq 0$  となる  $[\sigma] \in H_2(\mathcal{R}, \mathcal{F})$  が存在するならば、  
センサーは領域を被覆している



昔のGhrist



今のGhrist

最近は web lecture でもとても有名な数学者

# データに振り回されない

- ❖ Protein Data Bank (数十万個のタンパク質データ)
- ❖ MDシミュレーション(数万個の原子の運動)
- ❖ センサーネットワーク(数万個のセンサー情報)
- ❖ 株価予測(500万件/sec)
- ❖ 気象予測 etc.

どう理解するか

**粗視化と可視化 (幾何学) !!**

AYASDI (位相的データ解析ベンチャー企業)

---

# 数学って、, Abstract Nonsense だよね

---

## 社会的には

- ❖ 自分の研究にも取り込みたい
- ❖ 科学技術の基本言語

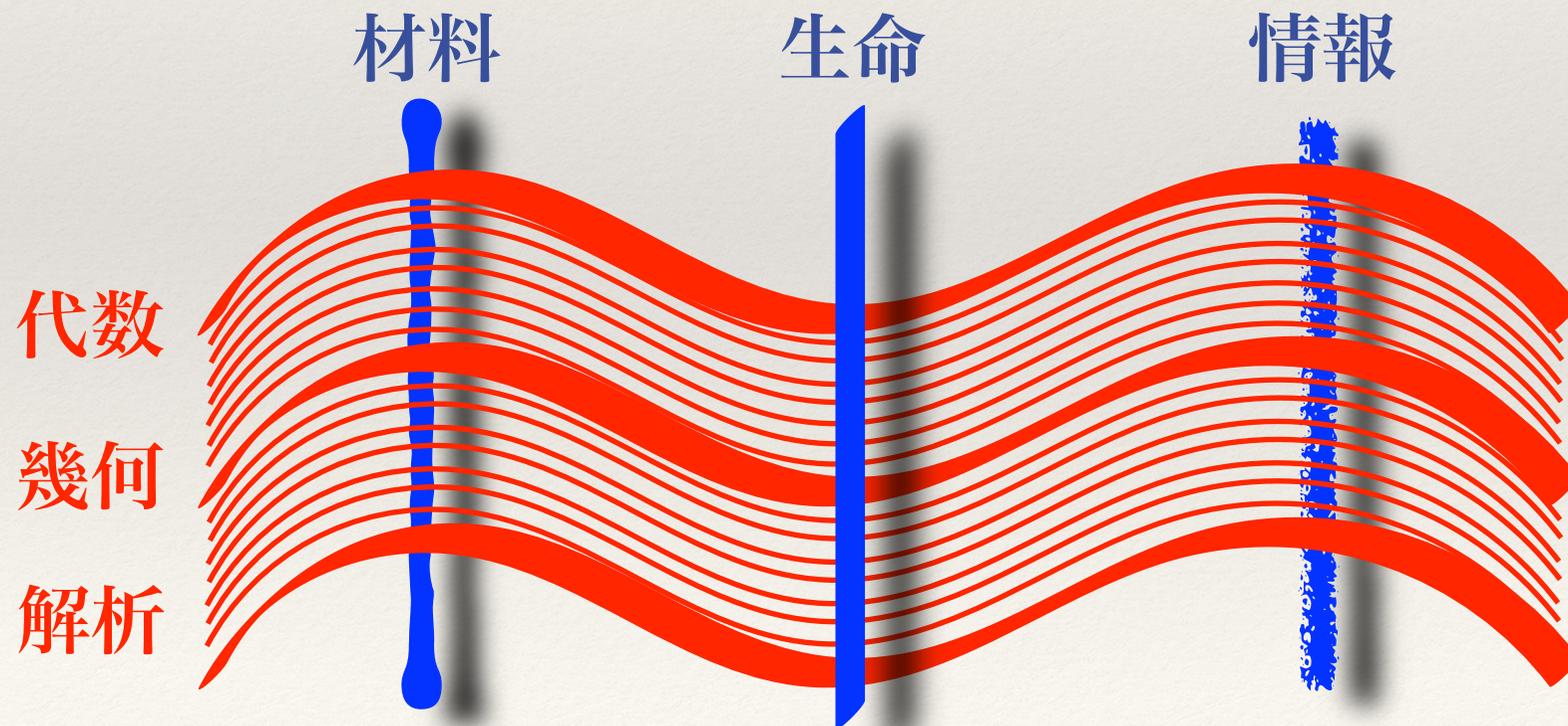
## でも本音は

- ❖ 不必要に抽象的
- ❖ 結局使えない

# Mathematics is Abstract Fullsense !!!

~~Nonsense~~

- ❖ 抽象性からくる普遍性 (横糸)
- ❖ 抽象性からくる創造性



## まとめ

1. やわらかく穴を扱う数学の道具としてホモロジー群がある
  2. 定義は少し難しかったが境界の特徴付けが大事そうだ
  3. いろいろな分野に応用されつつあるらしい
- 

数学では「定義」がとても重要です。興味を持った方は一度ホモロジー群の定義を自分で納得がいくまで考えてみてください。