

# 純粹数学者のための データサイエンス入門

株式会社アトラエ 杉山聡  
Alcia Solid



## 数学者 in DS とは？

数学者: 数学を分かっていて、使える人

DS: データを使って価値を生む人・チーム

数学者 in DS は以下の2種

- シンプルな数学を深く使って価値創出
- 深い数学をシンプルに使う価値創出

# 自己紹介

## 杉山 聡 (すぎやま さとし / 33)



2017年 3月  
2016年10月

博士(数理科学)取得  
Atrae, inc. 入社

→ マーケティングに従事

2018年10月  
現在

Data Scientist へ転向  
10人くらいの DS team に

### マーケ時代:

web 広告分析、機能企画、ユーザーヒアリング

### DS 時代:

数理統計中心(ベイズ x 微積分) + 機械学習  
データサイエンティスト協会 スキル定義委員  
[『分析モデル入門』](#)(ソシム、2022)

# 自己紹介 - 研究編

## 杉山 聡 (すぎやま さとし / 34)



### 研究テーマ：数学(数理解物理)

深谷圏の導来圏を用いて、有限次元代数の表現論における新しい計算公式を導出

### 研究内容が活きていること

なし！（ブラックホールの中に飛び込むビジネスが待たれる）

### 研究筋肉が活きていること

- 線形代数と微積分は100%理解してるYO！
- 数式の領域で創造性を発揮する経験
- 複雑なことを複雑なまま理解して処理

# 自己紹介(本題)

## アイシア=ソリッド (5)



2018年5月30日 生まれ

バーチャルデータサイエンティストとして  
YouTube を中心に活動中！



←登録してね！

いろんな動画を投稿してるよ！



いろんな動画を投稿してるよ！

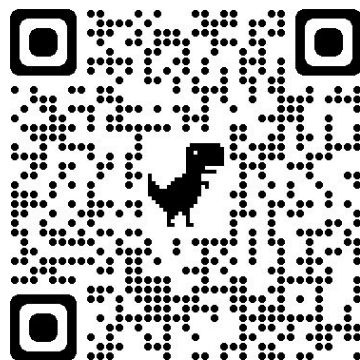


# 自己紹介(本題)

インターン募集してるよ！



- 動画編集(分析モデル系)
- 動画撮影・編集(インタビュー動画)
- YouTube shorts 作成
- 英語字幕作成スクリプト開発
- 分析モデル解説ブログ執筆・運営



←応募してね！  
(公式サイト Contact)

<https://speakerdeck.com/atrae/join-us>

# 温度・スピン・ 固有ベクトル

# 1. 温度

# 1. 温度

## 目次

- A. 問: 変動を検知せよ
- B. Bayes と温度
- C. 項目反応理論
- D. 階層 Bayes
- E. グループ潜在得点と温度
- F. 参考文献

NOT

今日で理解

# トピックス紹介 with 参考文献

A. 問：変動を  
検知せよ

# A. 問: 変動を検知せよ

## 背景: アンケートサービスでの DS 課題

### 問: 変動を検知せよ

毎月1回、従業員にサーベイ(アンケート)を送る  
回答から算出されたスコアの変化を検知したい

**ただし、単純にはいかない**

### 2種の平均

全32質問 → 26小項目 → 9キードライバー → 1総合スコア  
部署のスコアはメンバーの平均

部署の人数、項目によって、平均に用いられるデータ数が異なる  
→ 数値の変化しやすさが異なる！

# A. 問: 変動を検知せよ

## 質問 → 小項目 → キードライバー → 総合スコア

### 質問→小項目

Q16: 上司は誠実な態度で対応してくれていますか？

Q17: 上司と仕事以外の事でも、気軽に話すことができますか？

この2質問の平均で、小項目「上司との関係」が定義される

### 小項目→キードライバー

小項目「上司との関係」「仕事仲間との関係」の平均で、  
キードライバー「人間関係」が定義される

前へ

5  
16

上司と仕事以外の事でも、気軽に話すことができますか？

### キードライバー→総合スコア

9つあるキードライバーのうちの8つを平均して「総合スコア」が定義される

気軽に話すことができない

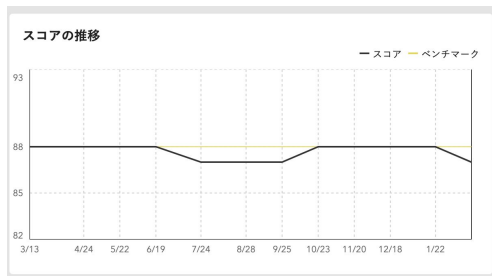
気軽に話すことができる



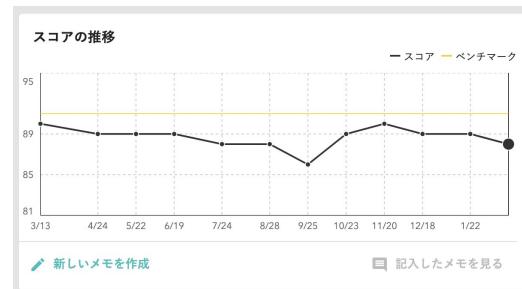
# A. 問: 変動を検知せよ

## スコアの変化のしやすさ

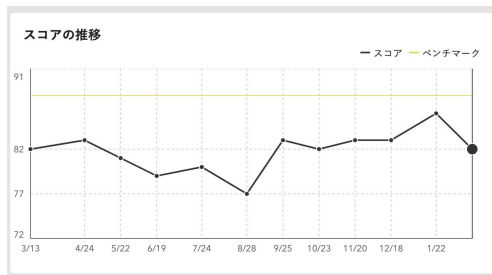
### 100人 - 総合スコア



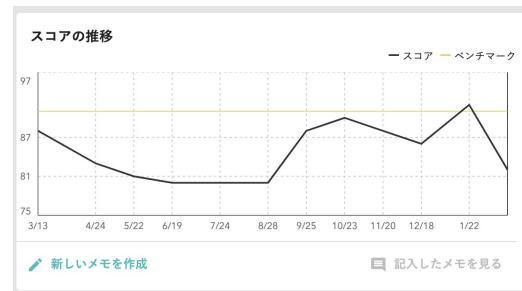
### 100人 - 小項目(やりがい)



### 5人 - 総合スコア



### 5人 - 小項目(やりがい)



# A. 問: 変動を検知せよ

p-値ではダメ  (検定知ってる人向け)

$n \rightarrow \infty$  で  $p \rightarrow 0$

実世界では帰無仮説 ( $X = 0$ ) が成立する可能性は0

データ数が増えれば、その差を必ず検出できる

実際、300人くらいを超えると、毎回変動が検知されるように

# A. 問: 変動を検知せよ

## 「温度回」の盛り上がりところ

変動を検知しよう！

そのために階層ベイズを使おう！

計算したら温度が出てきた！

そして、計算がめっちゃ高速化できた！（約1000倍）

# B. Bayes と温度

# B. Bayes と温度

## ベイズの定理

### ベイズの定理

$$P(X \mid Y) = P(Y \mid X) \times P(X) / P(Y)$$

$$(\text{prf: } P(Y \mid X) \times P(X) = P(X, Y) = P(X \mid Y) \times P(Y))$$

## B. Bayes と温度

### 最尤推定

問: コインを3回投げたら2回表だった → このコインが表になる確率は？  
表率を  $p$  とかく。

すると、2/3表になる確率  $L(p)$  は

$$L(p) = 3 \times p^2 \times (1-p)^1$$

$L(p)$  : 尤度 (Likelihood) という

この  $L(p)$  が最大になる  $p$  は  $p = 2/3$ 。

なので、このコインが表になる確率は  $p = 2/3$  !

**んなわけあるかい !**

## B. Bayes と温度

### 最尤推定の問題とベイズ化

#### 前の議論の問題点

表率  $p = 2/3$  であるコインの存在に疑問を挟まなかったこと  
 多少の歪みはあるとしても、 $50 \pm 5\%$  くらいでは？

#### ベイズ統計

「世の中にあるコインの表率は確率分布  $p^{\text{pre}}(p)$  に従って決まっている」と考える。

すると、表率が  $p$ 、かつ、 $2/3$  回表になる確率は、 $p^{\text{pre}}(p) \times L(p)$

例えば、 $p^{\text{pre}}(p) = C \times p^{50} \times (1-p)^{50}$  と設定すると (← 平均 50% 標準偏差 5%)

$$p^{\text{pre}}(p) \times L(p) = D \times p^{52} \times (1-p)^{51}$$

これが最大になる  $p$  は  $p = 52/103 = 50.49\%$  ← **まとも！**

## B. Bayes と温度

### ベイズ統計の考え方

#### 事前分布、尤度、事後分布

推定したいパラメーター $\theta$ が従う確率分布を考える  $p^{\text{pre}}(\theta)$

そのパラメーターで、観測データが得られる確率を計算  $p(X | \theta) = L(\theta)$

ベイズの定理を用いて、 $p(\theta | X)$  を計算

$$p^{\text{post}}(\theta) = p(\theta | X) = p(X | \theta) \times p^{\text{pre}}(\theta) / p(X) = L(\theta) \times p^{\text{pre}}(\theta) / p(X)$$

$p^{\text{post}}(\theta) = p(\theta | X)$  : データ  $X$  が得られた場合の、パラメーター $\theta$ の確率分布

$p^{\text{pre}}(\theta)$  を事前分布といい、 $L(\theta)$  を尤度といい、 $p^{\text{post}}(\theta)$  を事後分布という。

**事前分布を設定し、尤度と合わせ、事後分布を計算し、  
より正確な値を推定。**

# B. Bayes と温度

## Bayes と温度

ベイズの定理の式を微修正する

$$p^{\text{post}}(\theta) = L(\theta) \times p^{\text{pre}}(\theta) / p(X) \\ (p(X) = \int L(\theta) \times p^{\text{pre}}(\theta) d\theta)$$

↓逆温度 $\beta$ を導入 ( $\beta = 1/T$ )

$$p^{\text{post}}(\theta) = L(\theta)^\beta \times p^{\text{pre}}(\theta) / Z(\beta) \\ (Z(\beta) = \int L(\theta)^\beta \times p^{\text{pre}}(\theta) d\theta)$$

$\beta = 1$  : 普通のベイズ

$\beta \doteq 0$  :  $L(\theta)^\beta$  はほぼ1であり、データの影響が小さく、事前分布の影響が強い

$\beta \doteq \infty$  :  $L(\theta)^\beta$  の大小が与える影響が強く、事前分布の影響が弱い

# C. 項目反応理論

# C. 項目反応理論

## テストで項目を測る

普通のテスト: 正解率で受験者を評価する  
項目反応理論: 正解率で、問題も評価する

優秀 !

難問 !

|    | A氏 | B氏 | C氏 | D氏 | E氏 | F氏 |
|----|----|----|----|----|----|----|
| 問1 | ○  | ×  | ×  | ×  | ×  | ×  |
| 問2 | ○  | ○  | ×  | ×  | ×  | ×  |
| 問3 | ○  | ○  | ○  | ×  | ×  | ×  |
| 問4 | ○  | ○  | ○  | ○  | ×  | ×  |
| 問5 | ○  | ○  | ○  | ○  | ○  | ×  |
| 問6 | ○  | ○  | ○  | ○  | ○  | ○  |

# C. 項目反応理論

## ベイズを用いて平均点 → 潜在得点

### 普通の考え方

7段階で回答してもらった。

左から1点、2点、、、7点とし、平均点を計算

### 項目反応理論 x ベイズ

エンゲージメントの状態を表す潜在変数  $\theta \in \mathbb{R}$

$\theta$ は  $p^{\text{pre}}(\theta)$  に従う

$\theta$ に応じて、 $p_1(\theta), p_2(\theta), \dots, p_7(\theta)$  が定まる  
(項目反応関数 Item Response Function)

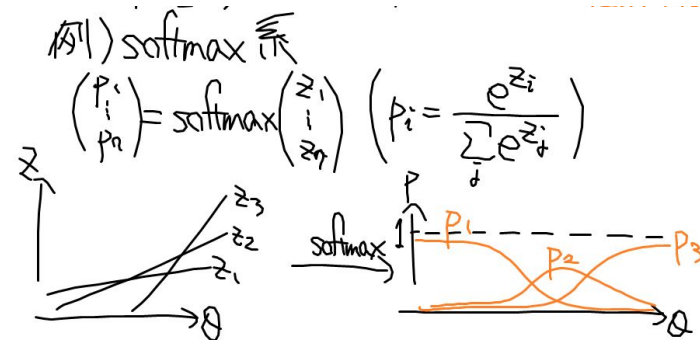
$p^{\text{pre}}(\theta)$  と  $p_i(\theta)$  をあわせて  
事後分布  $p^{\text{post}}(\theta)$  を計算し、  
これをもとにスコアを算出

前へ

5  
16

上司と仕事以外の事でも、気軽に話することができますか？

◀ 気軽に話することができない • 気軽に話することができる ▶



# D. 階層 Bayes

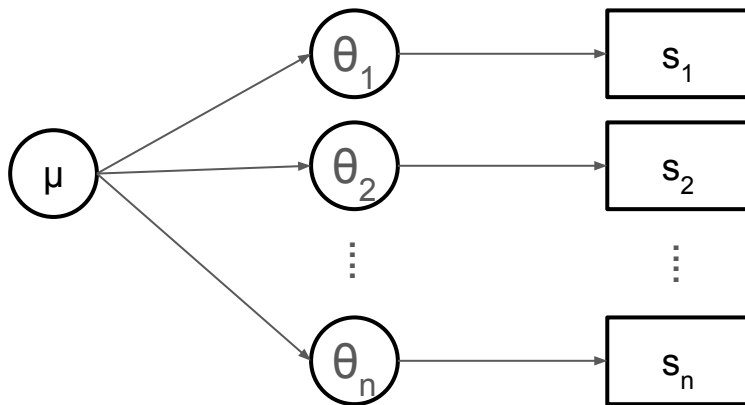
## D. 階層 Bayes

### 潜在変数に階層構造をもたせる分析手法

今回の問:

グループのスコアの変動を検知せよ!

階層ベイズモデルの例:



$\mu$ : グループの状態の潜在変数  
 $\theta_i$ : 従業員*i*の状態の潜在変数  
 $s_i$ : 従業員*i*の回答(1~7)

## D. 階層 Bayes

### 確率モデル

今回の設定:

$\mu$  の事前分布は正規分布 /  $\theta_i$  は平均  $\mu$ 、精度  $\lambda$  の正規分布 /  $s_i$  は IRF に従う

$$p^{\text{post}} = \frac{1}{Z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\mu^2\right) \times \frac{b^a}{\Gamma(a)} \lambda^{a-1} \exp(-b\lambda) \\ \times \prod_i \sqrt{\frac{\lambda}{2\pi}} \exp\left(-\frac{1}{2}(\theta_i - \mu)^2\right) \times p_{s_i}(\theta_i)$$

) 潜在変数  
潜在変数  
~7)

# E. グループ潜在得と温度

# E. グループ潜在得と温度

## やりたいこと

### Given:

回答データ  $s_1, s_2, \dots, s_n$

### やること:

$p^{\text{post}}$  が最大となる  $\mu$  を求める  
with いろんな近似と積分

### 結果:

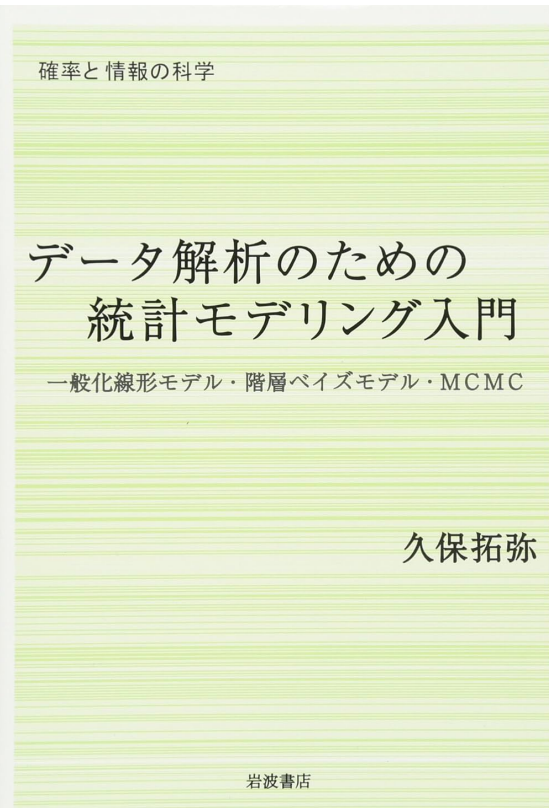
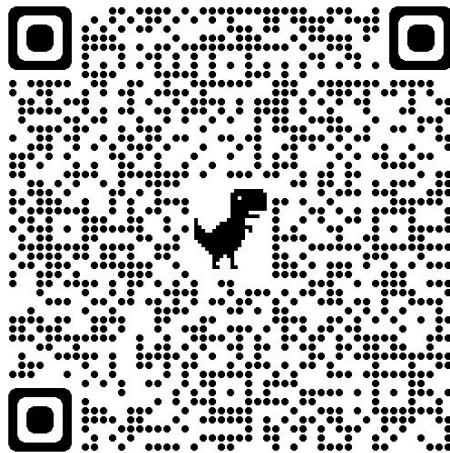
温度が出現する！  
計算が高速化できる！

# F. 参考文献

### 久保拓弥著 (2012)『データ解析のための統計モデリング入門』岩波書店

ベイズ統計の基礎から  
ベイズ統計を用いたモデリングまで

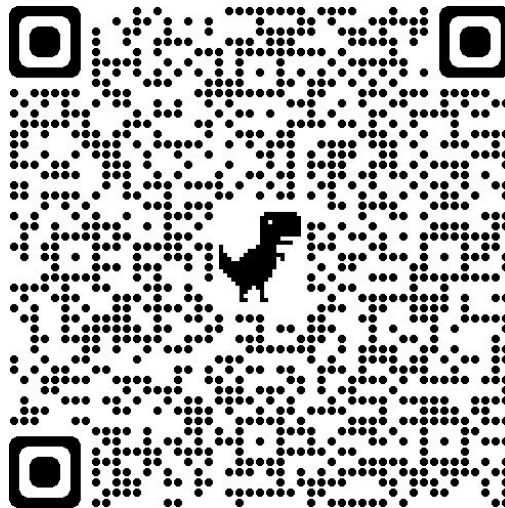
ベイズを用いたデータ分析の実践におすすめ！



## F. 参考文献

### 渡辺澄夫著 (2012)『ベイズ統計の理論と方法』コロナ社

ベイズ統計の数学的ラズボス本  
代数幾何の特異点解消を使って WAIC を定義する  
漸近理論を正確に記述している

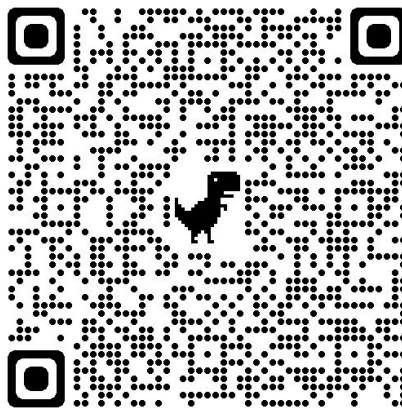


## F. 参考文献

### 豊田秀樹著 (2012)『項目反応理論[入門編](第2版)』朝倉書店

項目反応理論の入門と言えばこれ！

古典的テスト理論から、  
基礎的な項目反応理論まで紹介。  
他にも「中級編」「実践編」「理論編」とか  
いろいろあります



統計ライブラリー

### 項目反応理論 [入門編] 【第2版】

豊田秀樹  
[著]

朝倉書店

## F. 参考文献

### 実務に現れる数学 - 項目反応理論と温度について【第32回 数学カフェ】

今日の話をより詳しく紹介している動画です。  
イベントで4時間話したものがそのままアーカイブされています。



第 32 回 数学カフェ

#### データサイエンスの実務に現れる数学

項目反応理論と「温度」について



1 月 16 日 (土) 13 時から 17 時

Youtube チャンネル登録者数 1.5 万人超えの  
バーチャルデータサイエンティスト  
アイシア = ソリッドさんご講演

## F. 参考文献

「サンプルサイズが大きすぎると良くない」と思います

$n \rightarrow \infty$  で  $p \rightarrow 0$  の話を書いています。

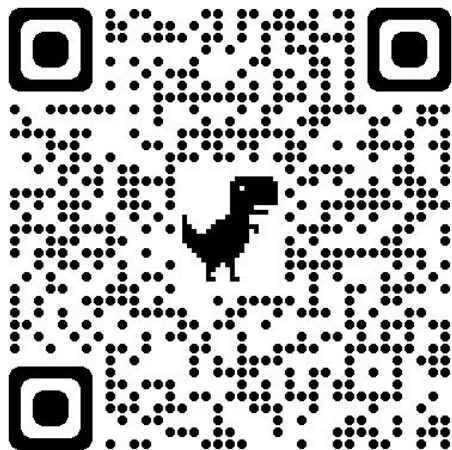


サンプルサイズ  
と  
p値の関係

# F. 参考文献

## 効果量(effect size)のはなし

$n \rightarrow \infty$  で  $p \rightarrow 0$  の問題に対処するには、「効果量」を用いるのが一般的です  
それについて一番わかり易いと思われるブログ記事



### 渋谷駅前で働くデータサイエンティストのブログ

元祖「六本木で働くデータサイエンティスト」です / 道玄坂→銀座→東京→六本木→渋谷駅前

#### 効果量(effect size)のはなし

R 統計学

2014-02-24

統計学的検定の話を始めたら自分の勉強の方が止まらなくなってしまったので（笑）、ついでにやってみようと思います。ちなみにこの記事は前回のやたらブックマ集めた記事の続きみたいなものです。

- そもそもビジネスの現場ではどういう「レベル」の統計学を使うべきなのか - 六本木で働くデータサイエンティストのブログ

例えば有意ではないという結果になった時にそれが「実際に帰無仮説が真」なのか「単にサンプルサイズが小さくて検出力が足りないだけ」なのか判断せよという問題。前者なら果てしなくサンプルサイズを大きくしても有意にはならないし、後者なら今度は効果量(effect size)のことを考えなければいけません。

#### プロフィール



TJO

Takashi J. OZAKI, Ph.D.  
Data Scientist (尾崎 隆)

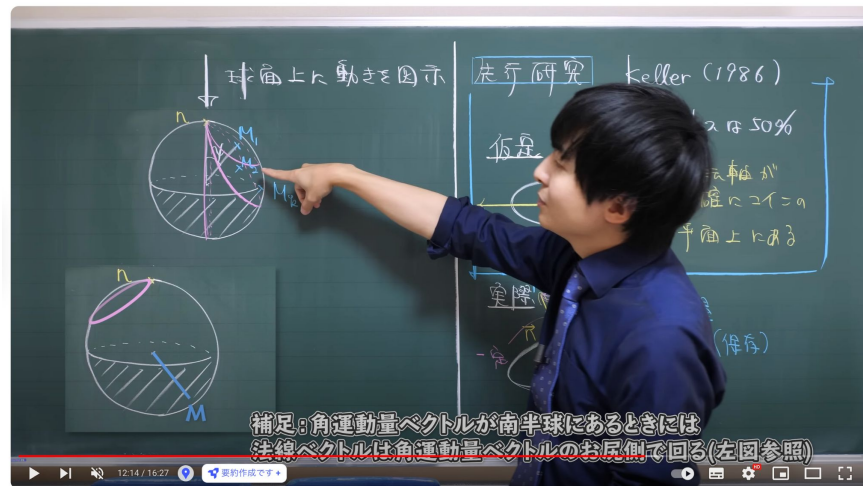
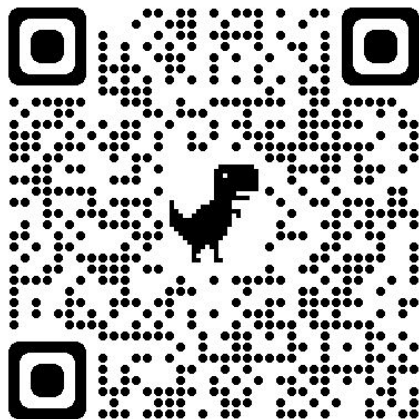
English: <https://tjo-hatenablog.com/>

このブログには Apache 2.0 ライセンスのもとで配布されている製作物が含まれています。

ブログの内容は個人の意見・見解の表明であり、所属組織の意見・見解を代表しません。またブログ内容の正確性

### コインスには偏りがある【物理論文解説】

本論とは関係ないけど面白いトピック  
コイン投げの表率は実は50%ではない(!)  
という論文を解説した動画です





# 温度・スピン・ 固有ベクトル

# 2. スピン

## 2. スピン

### 目次

- A. 問: 4次元物体を可視化せよ
- B. 被覆と Spin 群
- C. 4次元立体の回転の実装と Spin の自然な出現
- D. 参考文献

**A. 問: 4次元物体を  
可視化せよ**

# A. 問: 4次元物体を可視化せよ

まずは見て！

[https://drive.google.com/file/d/1iXOHfz5Y-oAqj5c9\\_BBu8xW5e2i4m8\\_d](https://drive.google.com/file/d/1iXOHfz5Y-oAqj5c9_BBu8xW5e2i4m8_d)

# A. 問: 4次元物体を可視化せよ

## 4Dロゴの作り方

### 2つのロゴの直積

会社のロゴ  $A \subset \mathbb{R}^2$  と、事業のロゴ  $B \subset \mathbb{R}^2$  の直積を取る  $\rightarrow A \times B \subset \mathbb{R}^4$   
 $\mathbb{R}^3 \subset \mathbb{R}^4$  との共通部分  $A \times B \cap \mathbb{R}^3 \subset \mathbb{R}^4$  を可視化

(ホントはロゴいじると  
怒られるけど、  
ノリで OK もらって  
ノリで作りました )

:Atrae  
wevox

# B. 被覆と Spin 群

## B. 被覆と Spin 群

### 3次元の回転を計算機で実現したい！

#### 四元数

$$r + xi + yj + zk \in H \ (r, x, y, z \in \mathbb{R})$$

$$i^2 = j^2 = k^2 = -1 \ / \ ij = k, \ jk = i, \ ki = j \ / \ ji = -k, \ kj = -i, \ ik = -j$$

#### 四元数を用いて回転を表現

$$\text{点 } P(x, y, z) \in \mathbb{R}^3$$

$$\rightarrow \alpha = 0 + xi + yj + zk \in H$$

$$\eta = \cos(\theta/2) + i \sin(\theta/2) \text{ を用いて}$$

$$\eta \alpha \bar{\eta} \text{ を計算すると、 } 0 + x'i + y'j + z'k \text{ の形をしていて、}$$

点  $Q(x', y', z')$  は点  $P$  を  $x$  軸を中心に  $\theta$  回転した点である

## B. 被覆と Spin 群

### Spin(3) $\sim$ $S^3 \rightarrow$ SO(3)

すべての回転は四元数で2通りで書ける

$\eta \in H$  を長さ1の四元数とすると

$a \rightarrow \eta a \bar{\eta}$  が回転に対応する

$S^3 \rightarrow$  SO(3) が得られる。

また、 $\eta$  と  $-\eta$  が表す回転は同一であり、そうでなければ別の回転を表す

$S^3 \rightarrow$  SO(3) は2重被覆 (2:1の写像)

# C. 4次元立体の回転の 実装と Spin の自然な出現

# C. 4次元立体の回転の実装と Spin の自然な出現

$$\text{Spin}(4) \sim \mathbf{S}^3 \times \mathbf{S}^3 \rightarrow \text{SO}(4)$$

すべての4次元回転は、四元数2つの組で表現できる

点  $(w, x, y, z) \rightarrow \alpha = w + xi + yj + zk \in \mathbb{H}$

$\eta, \xi \in \mathbb{H}$  を長さ1の四元数とする。

$\alpha \rightarrow \eta \alpha \bar{\xi}$  が4次元回転に対応する

実は、これで全部の4次元回転が表現できる。

また、 $(\eta, \xi)$  と  $(-\eta, -\xi)$  が表す回転は同一であり、そうでなければ別の回転を表す

$\mathbf{S}^3 \times \mathbf{S}^3 \rightarrow \text{SO}(4)$  は2重被覆 (2:1の写像)

## C. 4次元立体の回転の実装と Spin の自然な出現

**VR 空間で4次元物体を回そう！**

まず見て！

<https://www.youtube.com/watch?v=B-3VbgYqxNU&t=8420s>

## C. 4次元立体の回転の実装と Spin の自然な出現

$$\text{Lie}(\mathbb{R}^3 \rtimes \text{SO}(3)) \sim \mathbb{R}^3 \times \mathbb{R}^3 \sim \text{Lie}(\mathbb{S}^3 \times \mathbb{S}^3) = \text{spin}(4)$$

### 概要！

コントローラーの動きは平行移動 + 回転

接ベクトル(速度ベクトル)は3次元 + 3次元

片方の3次元を四元数の1つに、片方の3次元をもう1つの四元数に対応

Spin(4)の接ベクトル(速度ベクトル)を get

まとめると、コントローラーの速度ベクトルを、4次元回転の速度ベクトルに変換  
速度変化から回転量を再計算(積分 / exp)して回転を実現

# C. 4次元立体の回転の実装と Spin の自然な出現

## Spin の出現 !

実装の中身を見る(ざっくり / 厳密な実装は違います)

コントローラーの $\theta$ 回転( $x$ 軸周り)

$$\eta = \cos(\theta/2) + i \sin(\theta/2), \xi=1$$

$\theta = 360^\circ$  まで動かすと、

$\eta = -1, \xi = 1 \rightarrow$  これが表す変換は、

$$a \rightarrow \eta a \bar{\xi} = -a$$

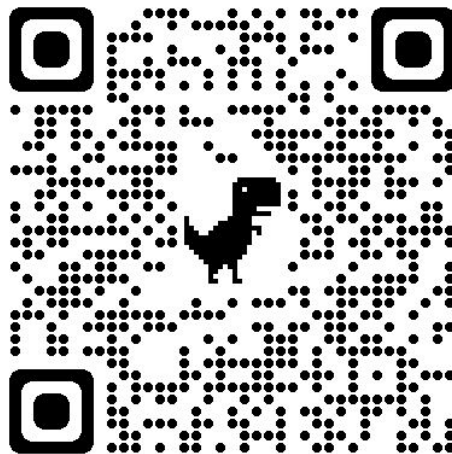
360°回転で-1倍だ !

# D. 参考文献

# C. 4次元立体の回転の実装と Spin の自然な出現

バーチャル学会はいいぞ！

VR 空間の「中で」やる学会  
2019から去年で5年目！  
ココで色々発表したものが元ネタです



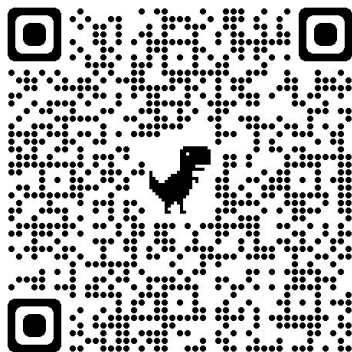
# C. 4次元立体の回転の実装と Spin の自然な出現

## 4次元立体の可視化 x 2

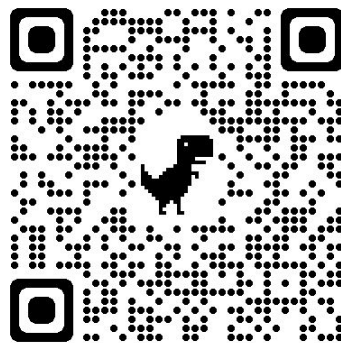
2021年に4次元VRロゴを実装・発表し  
2022年にその中に Spin を見た！ と発表



↓ロゴの回転動画



↓手で動かす動画(これ↑)



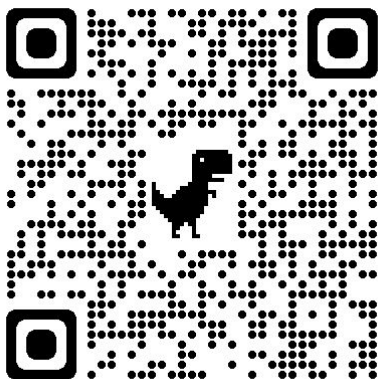
↓本発表の発表概要



# C. 4次元立体の回転の実装と Spin の自然な出現

## クォータニオンとは何ぞや？：基礎線形代数講座

四元数に関する、おそらく最もわかりやすい資料、というか教科書  
SEGA の方が書いて公開してくれている  
(ゲームでは3次元回転の高速計算が大事！)



# C. 4次元立体の回転の実装と Spin の自然な出現

## Spin について

頑張って勉強してください！

keyword:

トポロジーから攻めるなら「被覆」「基本群」あたり  
微分幾何から攻めるなら微分幾何の教科書を  
Lie群から攻めるならLie群の教科書を  
読みましょう！



# 温度・スピン・ 固有ベクトル

# 3. 固有 ベクトル

# 3. 固有ベクトル

## 目次

- A. 問:iさんに伝えるなら誰から話してもらう？
- B. グラフと中心性
- C. 情報伝達効率
- D. 参考文献

A. 問:iさんに伝えるなら  
誰から話してもらおう？

# A. 問:iさんに伝えるなら誰から話してもらう？

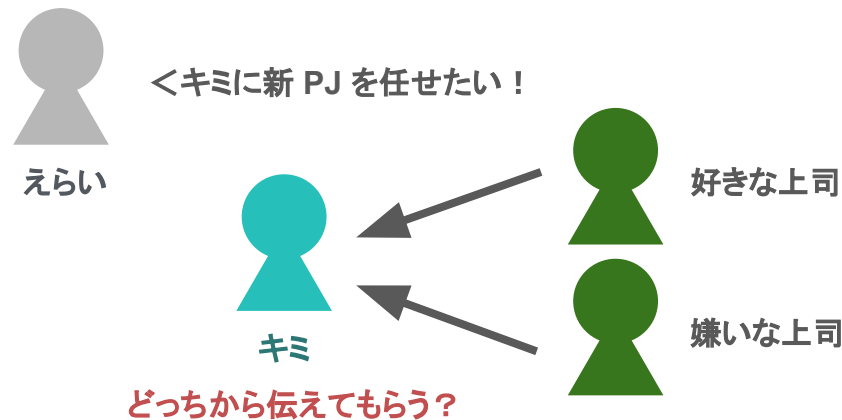
## 組織ネットワーク分析 (Organizational Network Analysis)

### ONA とは

組織(会社とか)の中のネットワークデータを分析して  
なんかいい感じになるようにする活動！

### 今回のテーマ

iさんに重大なことを伝えたい  
(「昇進だよ！」)  
誰から伝えるのが良いだろうか？



# A. 問:iさんに伝えるなら誰から話してもらう？

## 問題設定

### 大事なポイント

説得力ある人 ← network の中心性

キミの心に届く ← 距離が近い

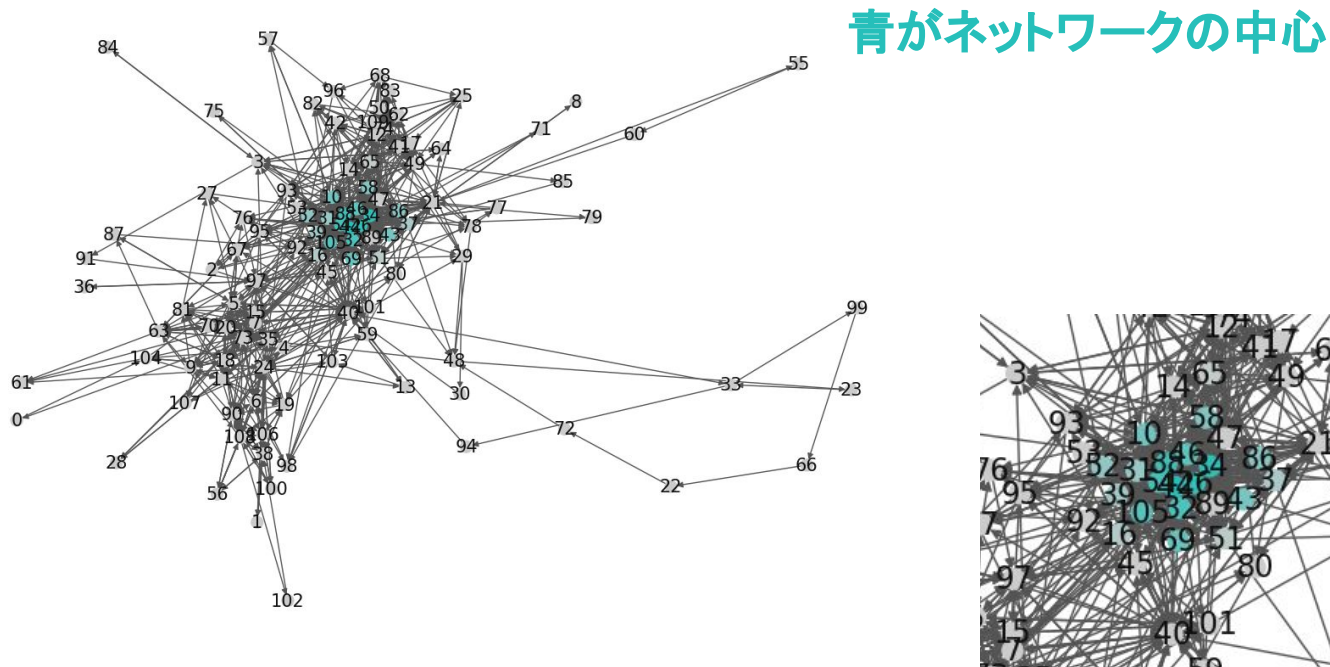
この2条件をいい感じに満たすアルゴリズムを使いたい！

### 今回のデータ

slack の会話データ

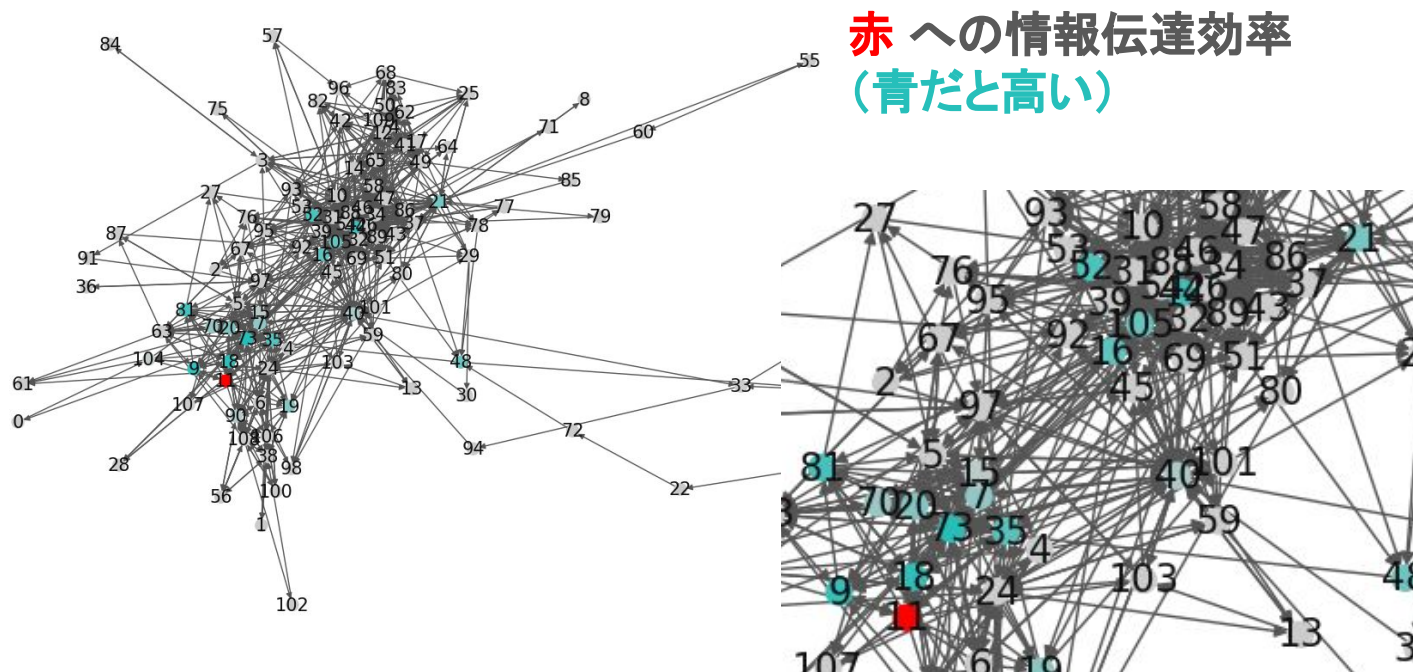
# A. 問:iさんに伝えるなら誰から話してもらおう？

## 先行研究:固有ベクトル中心性



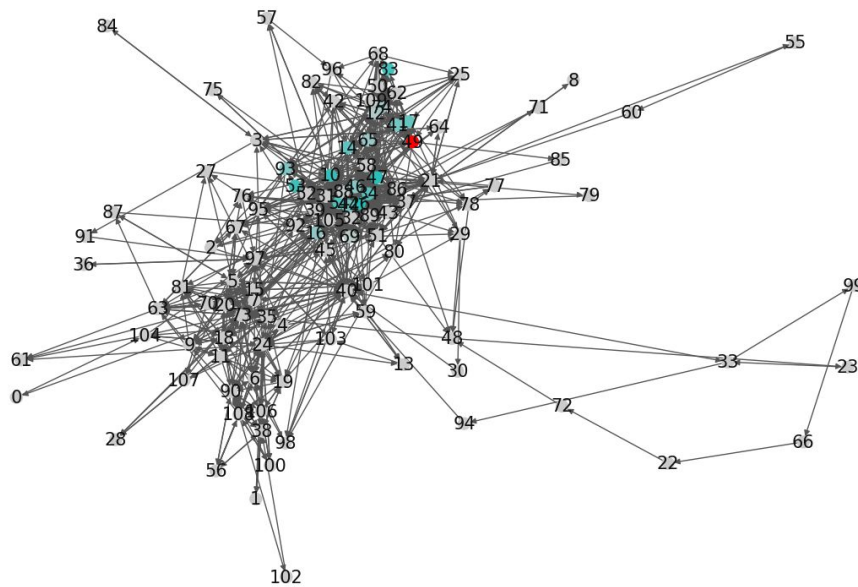
# A. 問:iさんに伝えるなら誰から話してもらおう？

## 今回の解①

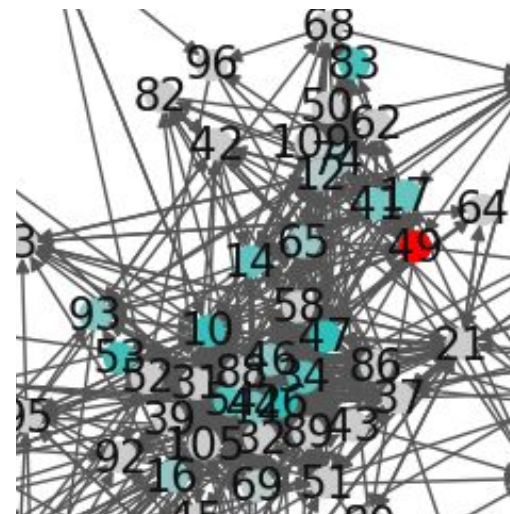


# A. 問:iさんに伝えるなら誰から話してもらおう？

## 今回の解②



赤 への情報伝達効率  
(青だと高い)



# B. グラフと中心性

# B. グラフと中心性

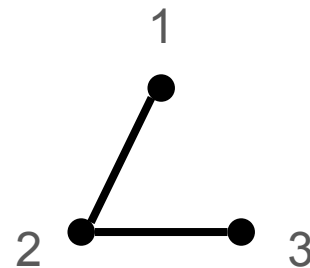
## グラフと有向グラフ

### グラフ

$G = (V, E)$  : graph

$V$ : set (of vertices)

$E \subset V \times V$ , s.t.  $(v_1, v_2) \in E \Rightarrow (v_2, v_1) \in E$ : set of edges



$V = \{1, 2, 3\}$

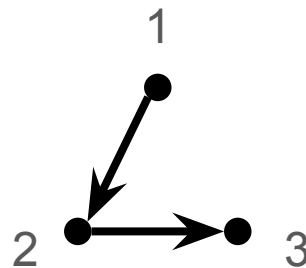
$E = \{(1, 2), (2, 1), (2, 3), (3, 2)\}$

### 有向グラフ

$G = (V, E)$  : directed graph

$V$ : set (of vertices)

$E \subset V \times V$ : set of edges



$V = \{1, 2, 3\}$

$E = \{(2, 1), (3, 2)\}$

# B. グラフと中心性

## グラフと有向グラフと隣接行列

### グラフと隣接行列

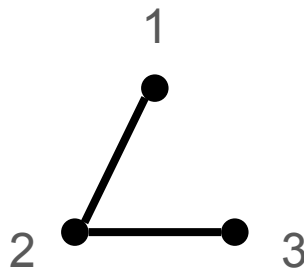
$G = (V, E)$  : (possibly directed) graph

$A = (a_{ij})$  : adjacency matrix

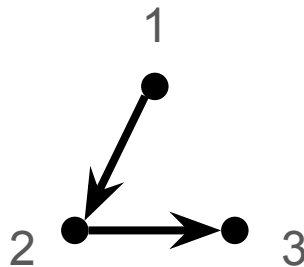
defined by

$a_{ij} = 1$  for  $(i, j) \in E$

$a_{ij} = 0$  otherwise



$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$



$$A = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

# B. グラフと中心性

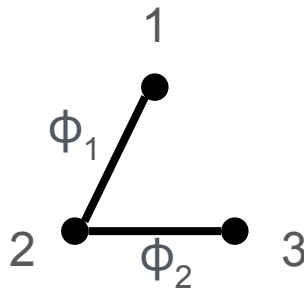
## 重み付きグラフと隣接行列

### 重み付き ver

$G = (V, E)$  : (possibly directed) graph

$\phi: E \rightarrow \mathbb{R}$  : weight

$(G, \phi)$ : weighted graph



$$A = \begin{bmatrix} 0 & \phi_1 & 0 \\ \phi_1 & 0 & \phi_2 \\ 0 & \phi_2 & 0 \end{bmatrix}$$

### 重み付きの隣接行列

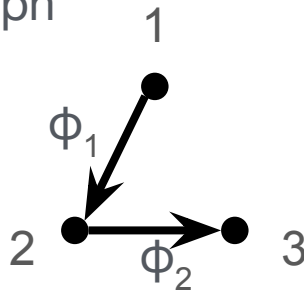
$(G, \phi)$  : weighted (maybe directed) graph

$A = (a_{ij})$  : weighted adjacency matrix

defined by

$a_{ij} = \phi((i, j))$  for  $(i, j) \in E$

$a_{ij} = 0$  otherwise



$$A = \begin{bmatrix} 0 & 0 & 0 \\ \phi_1 & 0 & 0 \\ 0 & \phi_2 & 0 \end{bmatrix}$$

## B. グラフと中心性

### いろんな中心性

#### 中心性とは

頂点の中心度合いを表す数値(やその計算手法)

#### 次数中心性

$G = (V, E)$ : graph

$v \in V$  の次数中心性  $\deg(v)$  は、

$\deg(v) = \#\{(v, -) \in E\} \text{ or } \#\{(-, v) \in E\} \text{ OR } \#\{(v, -) \in E\} + \#\{(-, v) \in E\} \dots$

$(G, \phi)$ : weighted graph の場合は

$\deg(v) = \sum_{e \in X} \phi(e)$

where  $X = \{(v, -) \in E\} \text{ or } \{(-, v) \in E\} \text{ OR } \{(v, -) \in E\} \cup \{(-, v) \in E\} \dots$

## B. グラフと中心性

### 固有ベクトル中心性

#### 固有ベクトル中心性とは

$G$  or  $(G, \phi)$  : (possibly weighted directed) graph

$A$ : its (possibly weighted) adjacency matrix

$w$ :  $A$  の最大固有値の固有ベクトル、長さ1、すべての成分が0以上のもの  
( $A$ の全成分が0以上の場合、(固有値に重複がなければ)一意に存在)

ベクトル  $w$  の第  $i$  成分  $w_i$  を頂点  $i$  の固有ベクトル中心性という

## B. グラフと中心性

### 固有ベクトル中心性②

#### 固有ベクトル中心性の性質

$w$ : 固有ベクトル中心性を与えるベクトル

$$w \propto Aw$$

$$(Aw)_i \propto \sum_j a_{ij} w_j$$

右辺は、 $i$ と隣接する or  $i \leftarrow j$  となる頂点  $j$  の中心性  $w_j$  の(重み付き)和

**固有ベクトル中心性は、中心にいるやつと隣接すると大きくなる**

# C. 情報伝達効率

# C. 情報伝達効率

## 問題設定再訪

問題:iさんに伝えるなら誰から？

説得力ある人 ← network の中心性

キミの心に届く ← 距離が近い

この2条件をいい感じに満たすアルゴリズムを使いたい！

### 単純な解法

説得力ある人 ← network の中心性 ← 固有ベクトル中心性？

キミの心に届く ← 距離が近い ← 隣接する頂点？

# C. 情報伝達効率

## 提案される情報伝達効率

jさんからiさんへの情報伝達効率

$w_j$ : 固有ベクトル中心性

$P$ : 遷移確率行列

$\lambda$ : 減衰パラメタ(分析者が指定)

$m$ : path length の上限パラメタ(分析者が指定)

$$\frac{w_j \left( \sum_{1 \leq k \leq m} \lambda^{k-1} P^k e_j \right)_i}{\sum_j w_j (\dots)_i}$$

# C. 情報伝達効率

## グラフの設定

今回利用するのは slack のデータ

$(G, \phi)$ : weighted directed graph generated by slack data

$G = (V, E)$

$V$ : 従業員の集合

$E$ :  $j$ さんから $i$ さんへ会話があれば  $(i, j) \in E$  (色々細かい設定あり)

$\phi((i, j))$ :  $j$ さんから $i$ さんへの会話の回数 (色々細かい設定あり)

## 確率遷移行列

$A = (a_{ij})$ : weighted adjacency matrix

$P = (p_{ij})$ : transition probability matrix (遷移確率行列)

defined by  $p_{ij} = a_{ij} / \sum_i a_{ij}$  (列の成分和が1になるように調整)

# C. 情報伝達効率

## 情報伝達効率 re-visited

j の中心性

$\lambda$  で減衰

k 次の隣接

$$w_j \left( \sum_{1 \leq k \leq m} \lambda^{k-1} P^k e_j \right)_i$$

おすすめ

$\lambda = 0.5$

$m = 3$

$$\sum_j w_j (\dots)_i$$

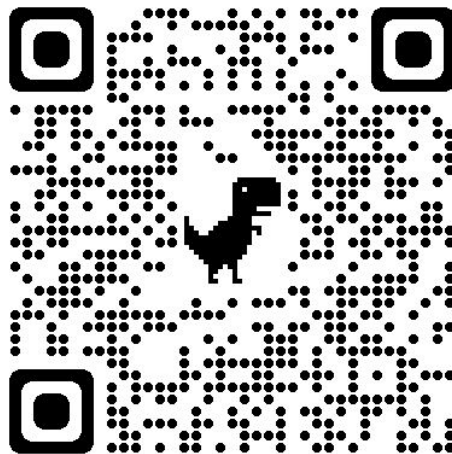
j から i

# D. 参考文献

## D. 参考文献

### バーチャル学会はいいぞ！

VR 空間の「中で」やる学会  
2019から去年で5年目！  
ココで色々発表したものが元ネタです



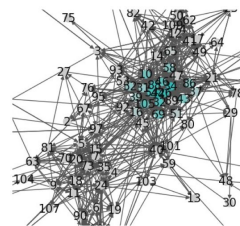
## 組織ネットワーク分析における新しい情報伝達効率の定義とその理論解析

バーチャル学会での発表内容の概要がこちら



組織ネットワーク分析における  
新しい情報伝達効率の定義とその理論解析

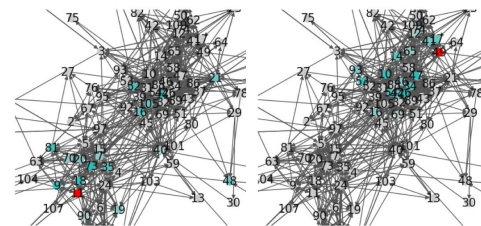
$$\text{j から i への情報伝達効率} := \frac{w_j \left( \sum_{1 \leq k \leq m} \lambda^{k-1} P^k e_j \right)_i}{\sum_j w_j (\dots)_i}$$



中心性



個別化

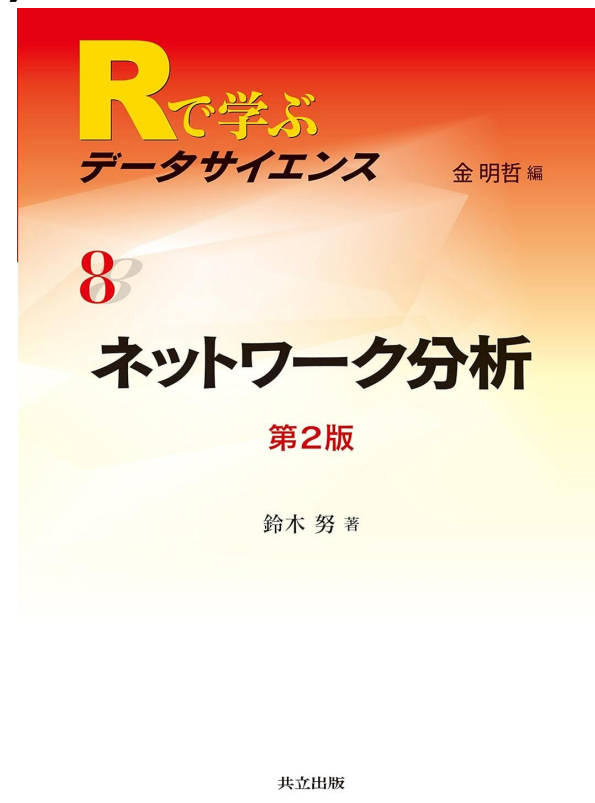
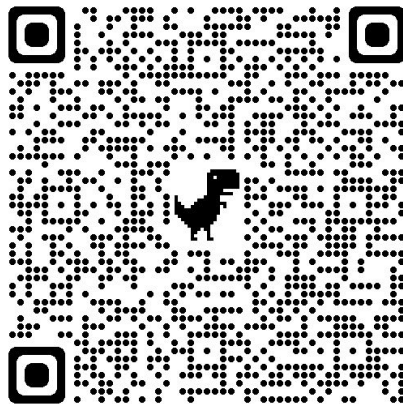


対象（赤）別の情報伝達効率上位（青）

## D. 参考文献

### ネットワーク分析 第2版 (Rで学ぶデータサイエンス 8)

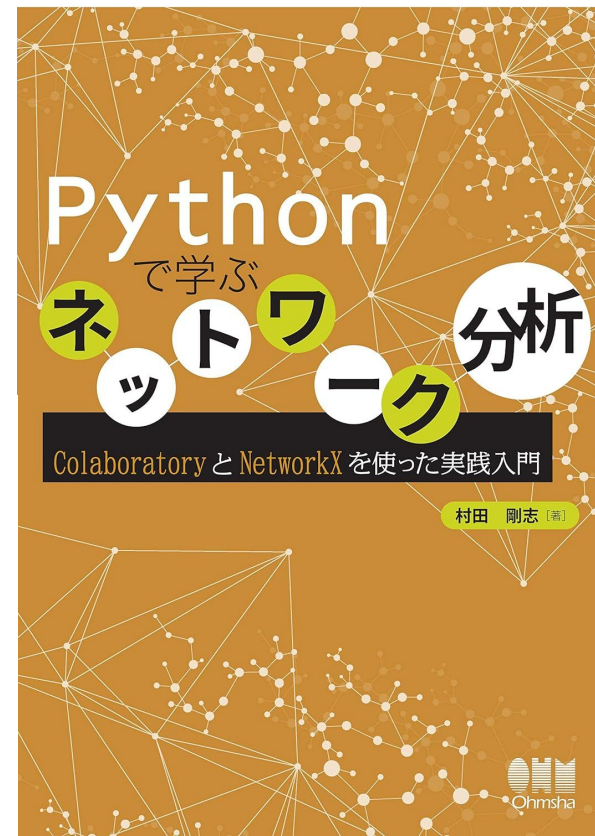
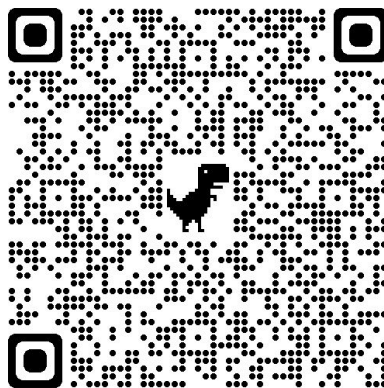
ネットワーク分析の基礎はこれがおすすめ！  
統計的技法がたくさん！  
(一般に、R を使う本のほうが、  
Python を使う本より、技法がリッチ。)



## D. 参考文献

### Pythonで学ぶネットワーク分析

Python で触りたい！  
という人にはこちらがおすすめ





Title

Subtitle  
content

久保拓弥著 (2012) 『データ解析のための統計モデリング入門』岩波書店

Chartr. (2022, December 9). ChatGPT: The AI Bot Taking the Tech World by Storm. <https://www.chartr.co/stories/2022-12-09-1-chatgpt-taking-the-tech-world-by-storm> (Accessed: 20XX, XXX X)

SpaceX. (2021, March 17). Starship | SN10 | High-Altitude Flight Recap [Video]. YouTube. <https://www.youtube.com/watch?v=gA6ppby3JC8> (Accessed: 20XX, XXX X)

やなぎ @yanagi\_aiai. (2023, February, 12). Twitter. [https://twitter.com/Yanagi\\_aiai/status/1624434539199356928?t=zsVEFPqbr3aT\\_9BwhowSDw&s=19](https://twitter.com/Yanagi_aiai/status/1624434539199356928?t=zsVEFPqbr3aT_9BwhowSDw&s=19) (Accessed: 20XX, XXX X)