

非バックトラックを活用した 複雑ネットワーク解析

小蔵 正輝

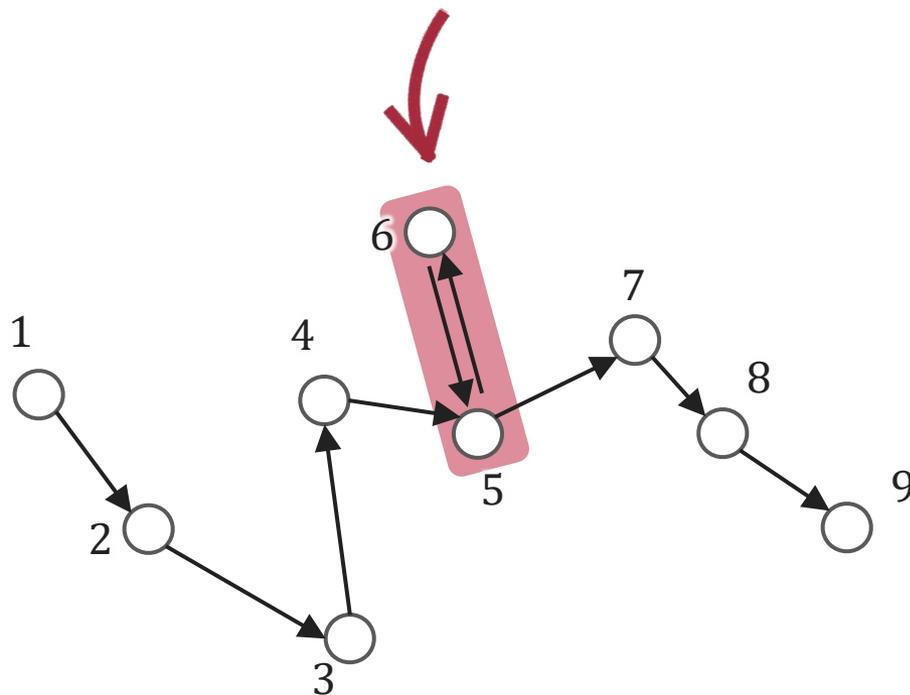
大阪大学
大学院情報科学研究科
バイオ情報工学専攻

はじめに

このような経験はありませんか

- シミュレーション用プログラムの手戻り
- お店に行ったが既に閉まっていた
- 送ったメールがエラーで返ってきた

それは**バックトラック**です



はじめに

本チュートリアルでは...

- 複雑ネットワークにおける非バックトラックの概念
- 非バックトラックの酔歩, 伝播, 中心性における応用
- 非バックトラック行列

(複雑)ネットワーク

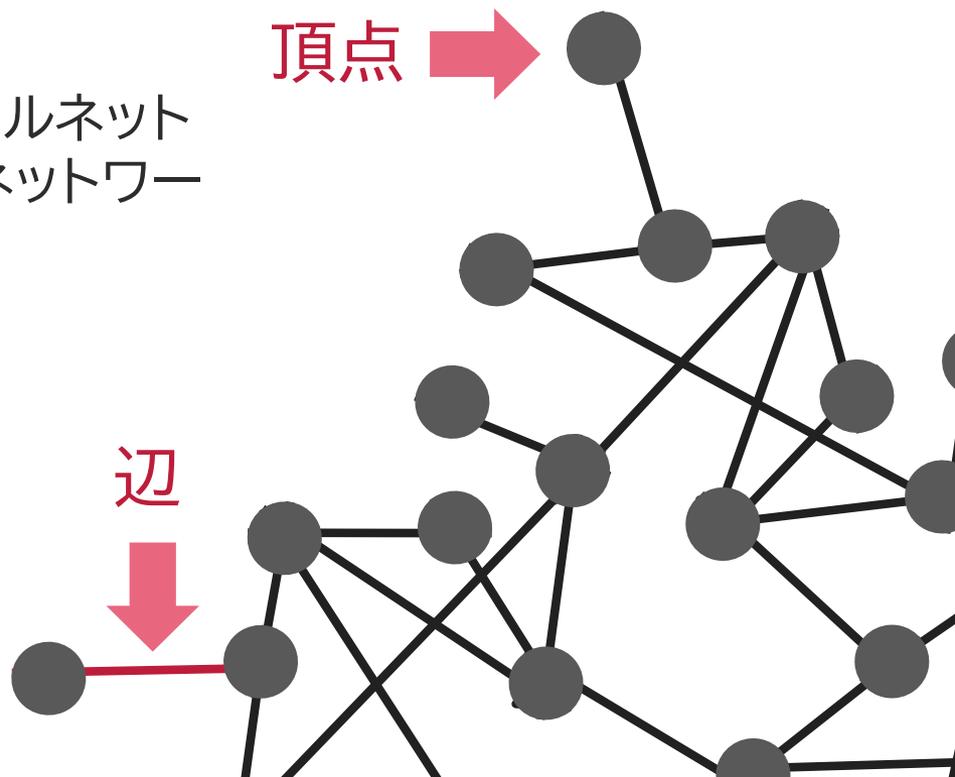
- グラフと同義. 時と状況において使い分ける.
- **頂点**: ネットワークにおける点のこと. **ノード**ともいう.
- **辺**: ネットワークにおける線のこと. **弦**, **弧**, **紐帯**, **リンク**とも.
 - **重要**: 任意の2つの頂点の間に複数の辺は**存在しない**と約束する
 - **重要**: 同じ頂点を結ぶ辺は**存在しない**と約束する

例

- 知人ネットワーク, オンラインソーシャルネットワーク, 科学技術ネットワーク, 文献ネットワーク, ...

ネットワーク科学の動機

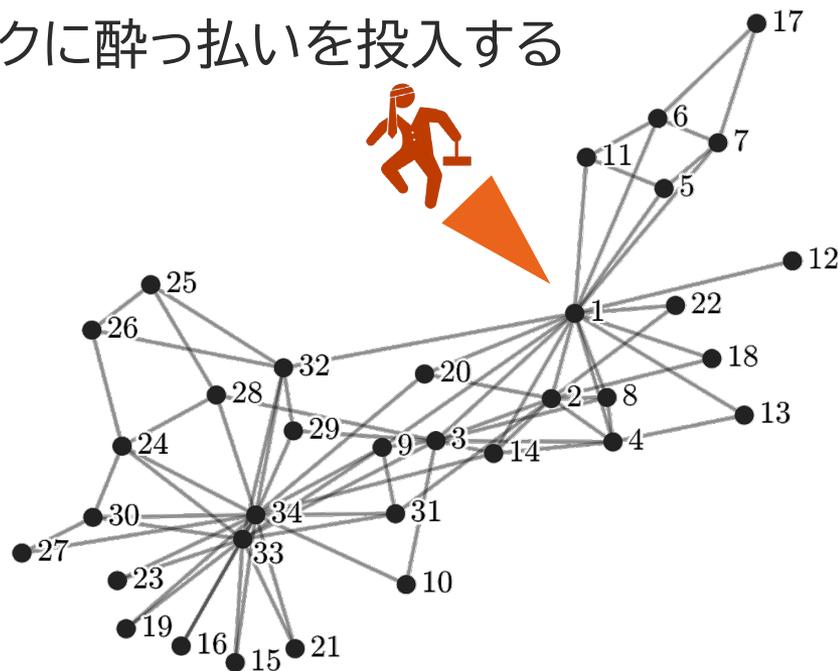
- 全体を正確に把握することが困難な大規模複雑ネットワークを解析するための方法論の確立



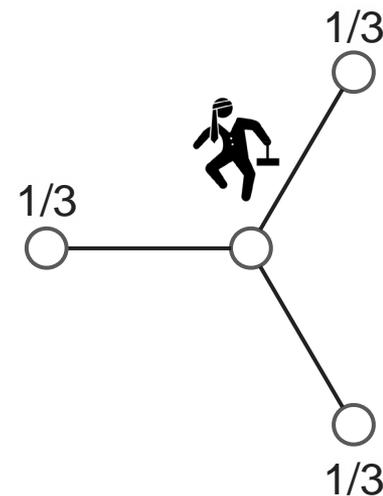
醉步

ネットワーク上の酔歩

複雑ネットワークに酔っ払いを投入する



動き方



- グラフの探索: 単純なルールを有するエージェントによる探索の実現
- 数学的な興味: 酔っ払いの位置の分布はどのように時間変化するか?
- 中心性の計算

酔っ払いは**バックトラック**するかもしれない

- (実物を見たことは無いが...)
- マルコフ過程
- グラフ探索の観点からは無駄に

問

- バックトラックを**禁止**するとグラフ探索の効率は向上するか？向上するならばどの程度か？



仮定

- すべての頂点は d 個の隣接頂点をもつ

定理

- 非バックトラック酔歩の定常分布への収束レートは

$$\tilde{\rho} = \phi\left(\frac{|\lambda|}{2\sqrt{d-1}}\right) / \sqrt{d-1} \quad \text{ただし } \phi(x) = \max(1, x + \sqrt{x^2 - 1})$$

- 参考: 通常の酔歩の収束レートは

$$\rho = |\lambda| / d$$

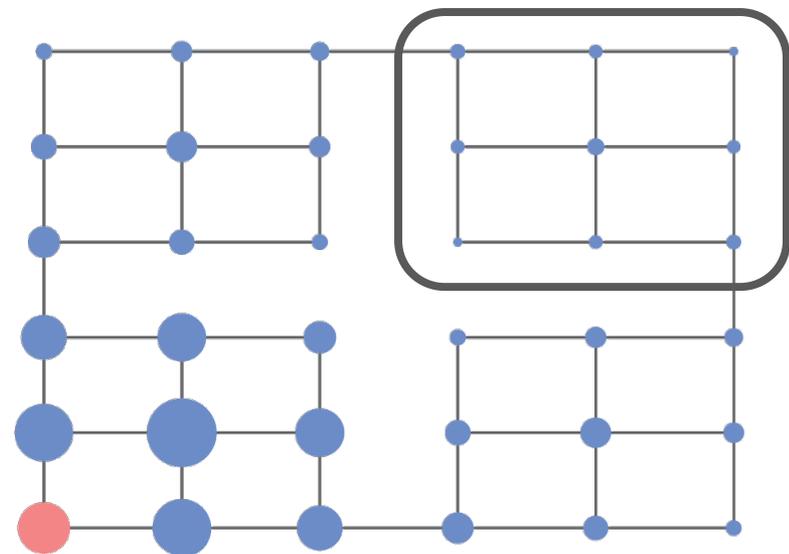
解釈

- 常に加速が実現: $\tilde{\rho} / \rho < 1$
- $|\lambda| = 2\sqrt{d-1}$ のときに最大の加速: $\tilde{\rho} / \rho = 1/2$

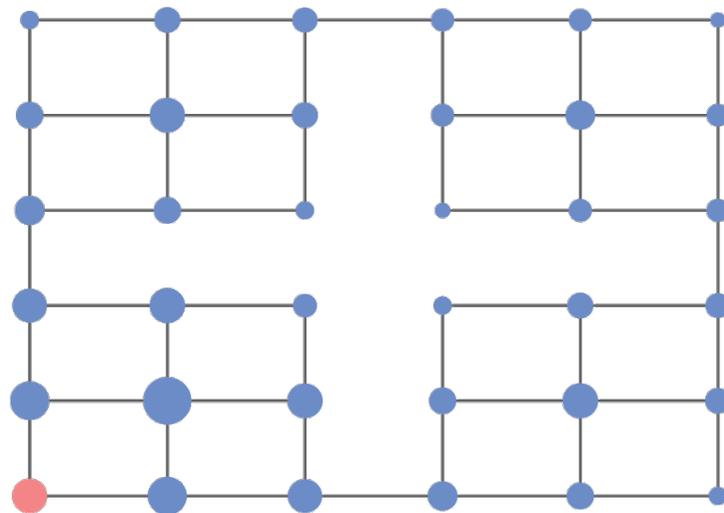
Alon, Benjamini, Lubetzky, and Sodin, "Non-backtracking random walks mix faster," *Commun. Contemp. Math.*, 2007.

- 四つの格子をつなげたネットワーク
- 時刻 $k = 1$ から $k = 36$ の酔歩における累積訪問回数の分布

うまく探索できていない領域



(a) 通常の酔歩



(b) 非バックトラック酔歩

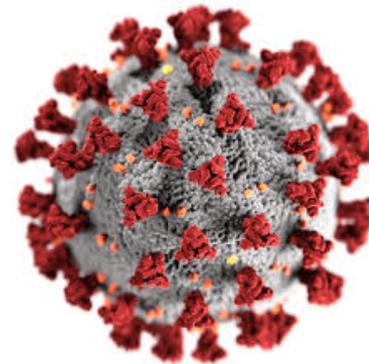
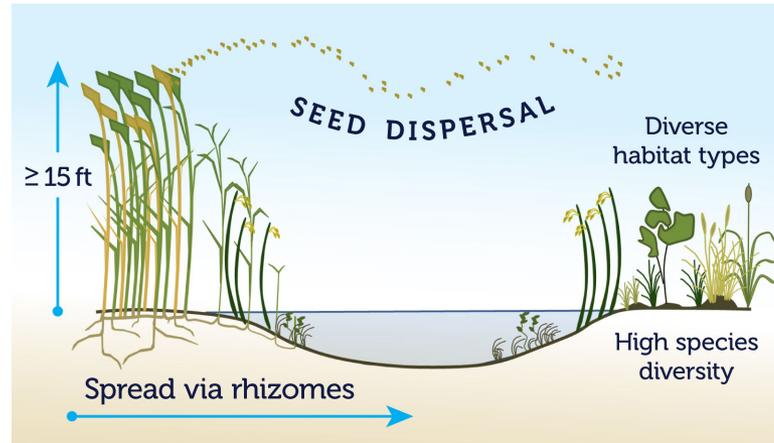
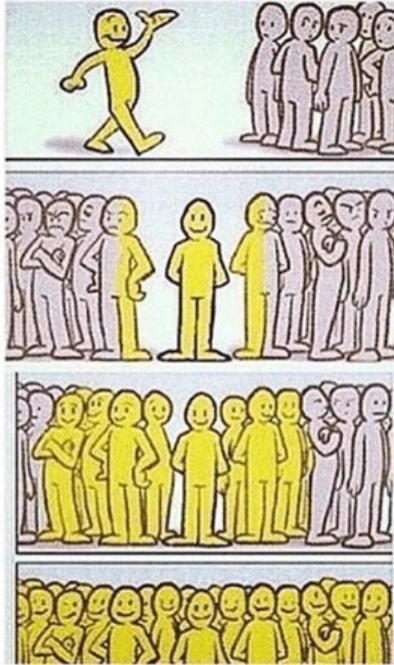
ネットワーク全体の
満遍ない探索

伝播

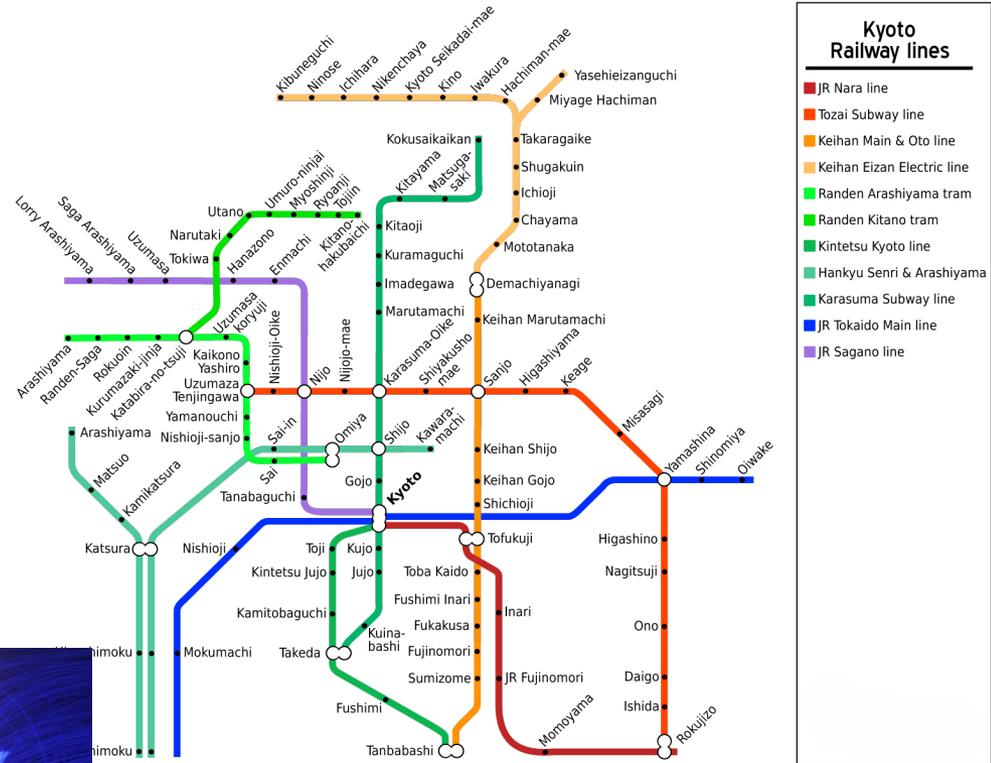
Masuda, Preciado, and Ogura, "Analysis of the susceptible-infected-susceptible epidemic dynamics in networks via the non-backtracking matrix," *IMA J. Appl. Math.*, 2020

Ogura, "Second-order moment closures for networked susceptible-infected-susceptible model," *52nd Int. Symp. Stoch. Syst. Theory Its Appl.*, 2020.

Let's spread Happiness and Good vibes..



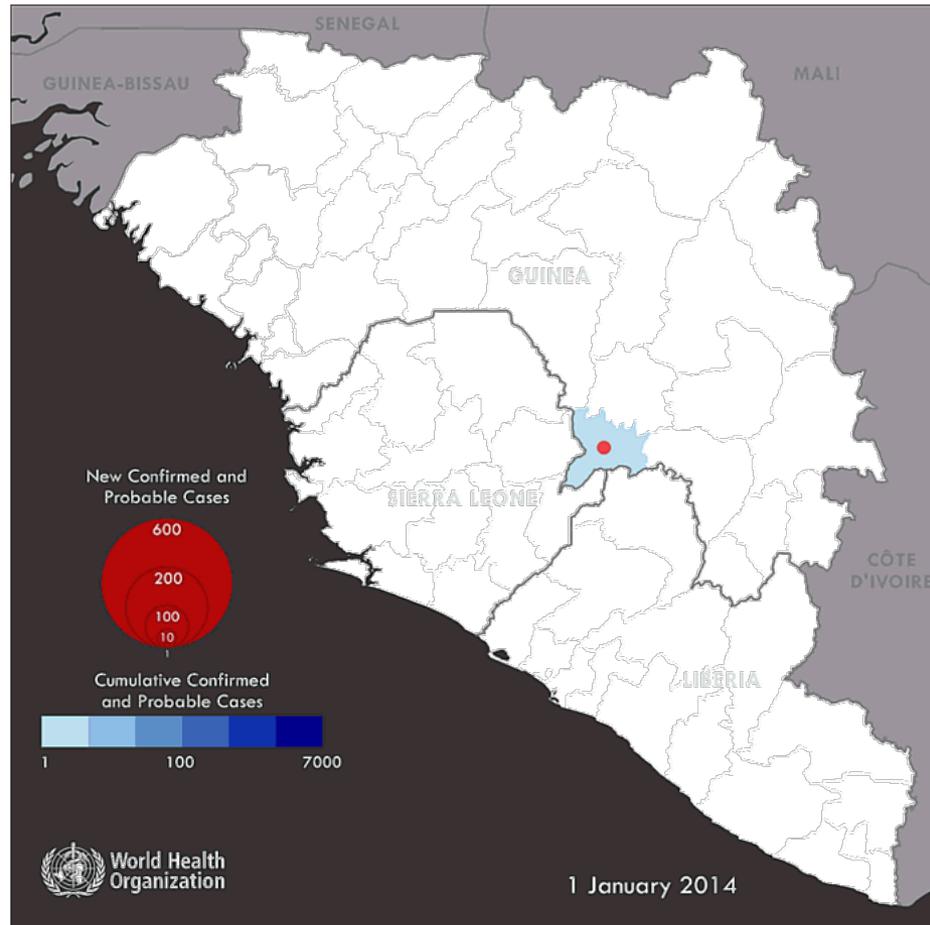
Happiness spreads :) Be a contributor, and tag a friend who needs to see this.



Information spread: 2011 Tohoku earthquake

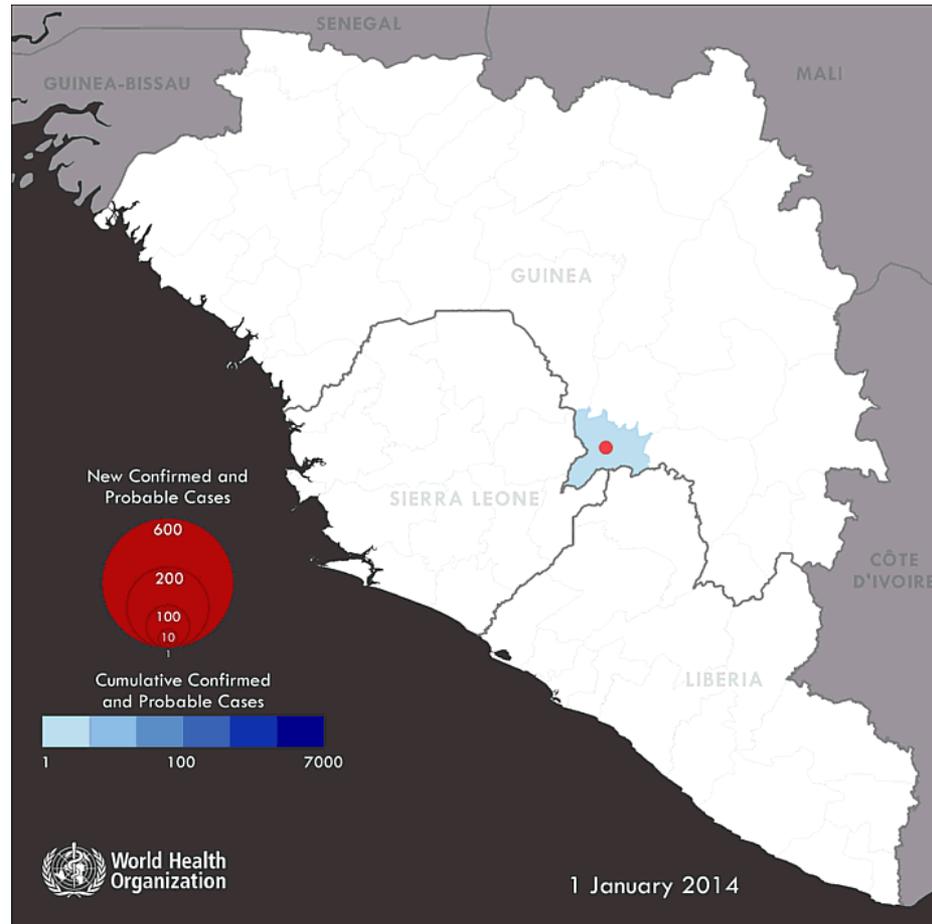
norbertomoreau, "Global pulse through Retweets", <http://en.inmoreau.com/720/global-pulse-through-retweets-2-video-grafics/>

Disease spread: Ebola outbreak



World Health Organization, <http://www.who.int/csr/disease/ebola/maps/en/>

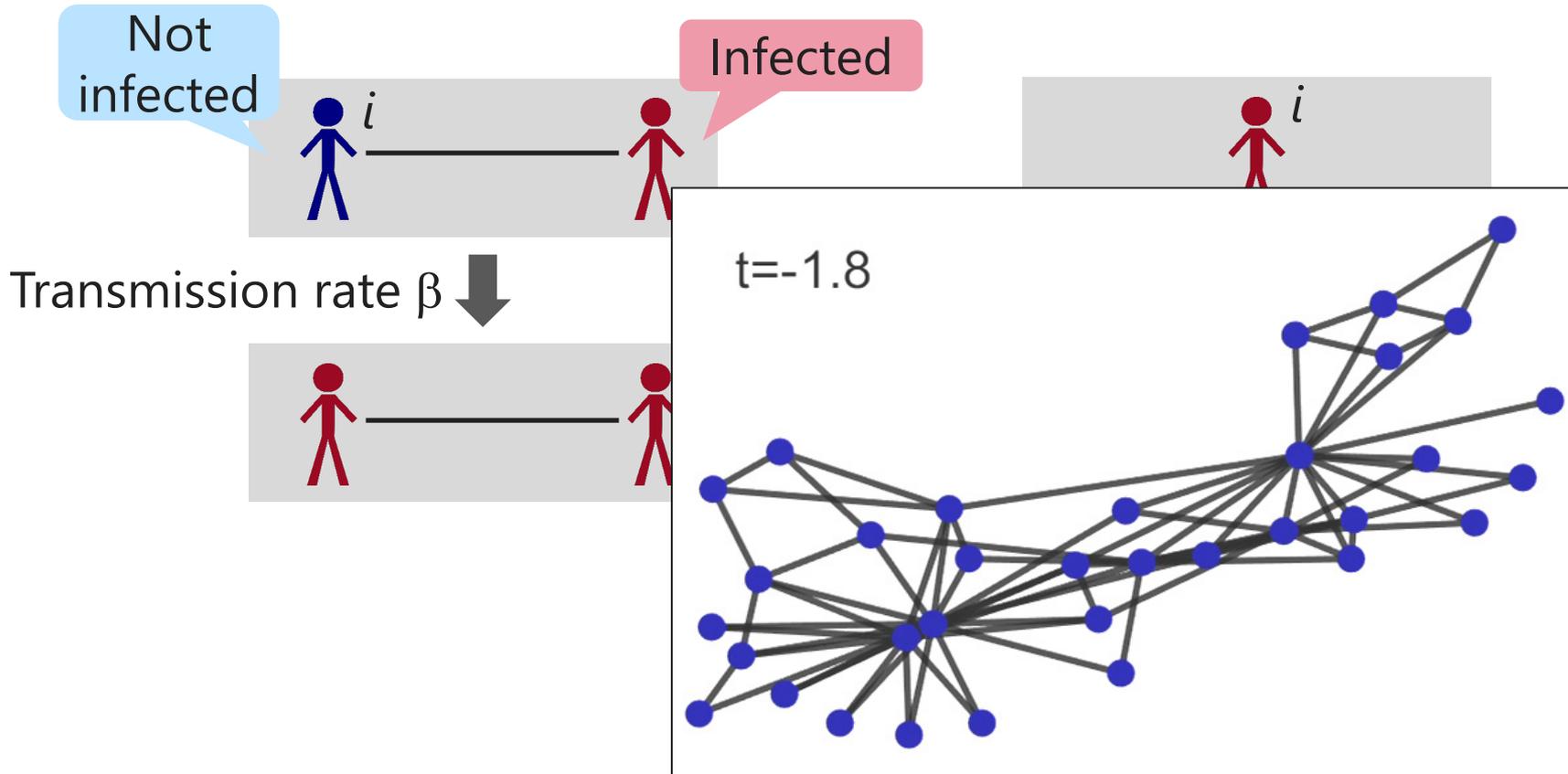
Disease spread: Ebola outbreak



World Health Organization, <http://www.who.int/csr/disease/ebola/maps/en/>

Networked model

Nodes = agents, edges = relationships



Pastor-Satorras, Castellano, Van Mieghem, and Vespignani, "Epidemic processes in complex networks," Rev. Mod. Phys., vol. 87, no. 3, pp. 925–979, 2015.

2^N -states Markov process

- $N = \#$ nodes (\geq hundreds, typically)
- 2 states per node
- Computationally infeasible

- Define

$$x_i(t) = \begin{cases} 1, & \text{if node } i \text{ is infected at time } t \\ 0, & \text{otherwise} \end{cases}$$

$$p_i(t) = \text{Prob}(\text{node } i \text{ is infected at time } t)$$

$$a_{ij} = \begin{cases} 1, & \text{if nodes } i \text{ and } j \text{ are adjacent} \\ 0, & \text{otherwise} \end{cases}$$

- Differential inequality for p

$$\frac{dp_i}{dt} = \beta \sum_{j=1}^N E[(1 - x_i)x_j] - \delta E[x_i]$$

$$\leq -\delta p_i + \beta \sum_{j=1}^N a_{ji} p_j$$

ここ: バックトラックを許して楽に解析
現実: 感染症はバックトラックしない

- Used in **most** of the papers on the control of SIS models!

- (Decay rate of $E[\|p(t)\|]) \geq \delta - \beta \rho(A)$

- Not necessarily accurate!

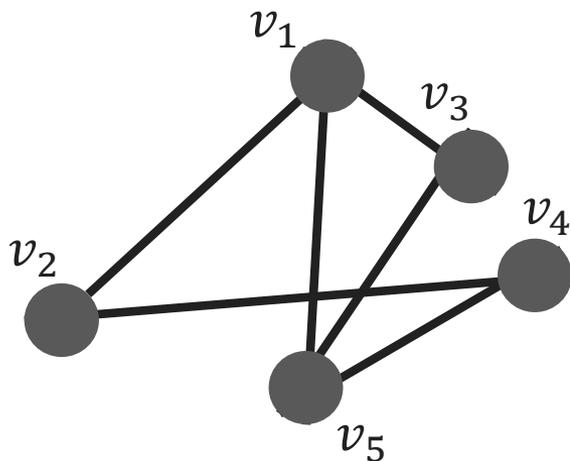
隣接行列

- 以下のように定められる行列 A をグラフの隣接行列と呼ぶ。

$$A_{ij} = \begin{cases} 1, & \text{頂点 } v_i \text{ と } v_j \text{ の間に辺がある場合} \\ 0, & \text{辺が無い場合} \end{cases}$$

- 計算機にとっては、グラフの絵よりも理解しやすい
- 数式でグラフを扱うときにも便利

- 例



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

- Second-order moments
- New variable

$$q_{ij}(t) = \text{Prob}(\text{nodes } i \text{ and } j \text{ are infected at time } t)$$

- A straightforward calculation yields...

$$\frac{dp_i}{dt} = \left(\sum_{j=1}^N E[(1 - x_i)x_j] \beta \right) - \delta E[x_i]$$

$$= \left(\sum_{j=1}^N \beta q_{ji} \right) - \delta p_i,$$

$$\frac{dq_{ij}}{dt} = - \left(\sum_{k=1}^N E[x_i(1 - x_j)x_k] \beta \right) + \delta E[x_i x_j]$$

$$+ \left(\sum_{k=1}^N E[(1 - x_i)(1 - x_j)x_k] \beta \right) - \delta E[x_i(1 - x_j)]$$

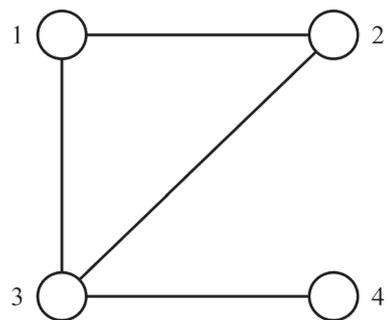
Want to express in terms of p or q !

Which is better?

- $E[(1 - x_i)(1 - x_j)x_k] \leq E[(1 - x_j)x_k]$.
 - [Ogura, Preciado, Systems and Control Letters, 2018]
 - Lower bound in terms of $N^2 \times N^2$ matrix

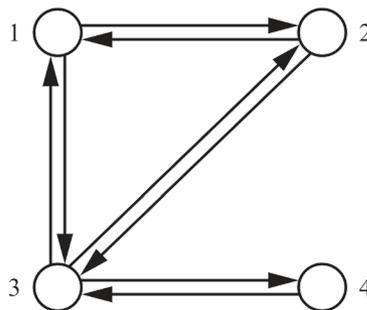
- $E[(1 - x_i)(1 - x_j)x_k] \leq E[(1 - x_i)x_k]$.
 - [Masuda, Preciado, Ogura, IMA Journal of Applied Mathematics, 2020]
 - Lower bound in terms of $(N + M) \times (N + M)$ matrix
 - M = number of edges
 - $N + M \ll N^2$ in many practical cases
 - There appears the **non-backtracking matrix**
 - **Yields better lower bound (not proven yet though)**

バックトラックではないパスを作る有向辺の組を関係付ける

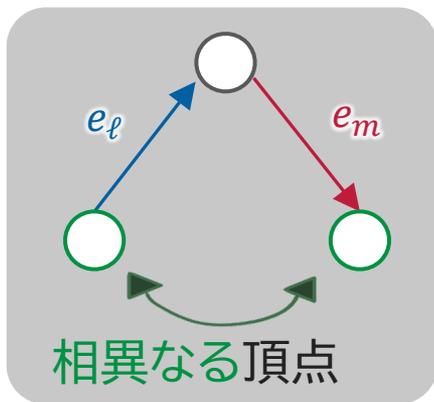


(a) 無向グラフ

辺を分解



(a) 有向グラフ D_G



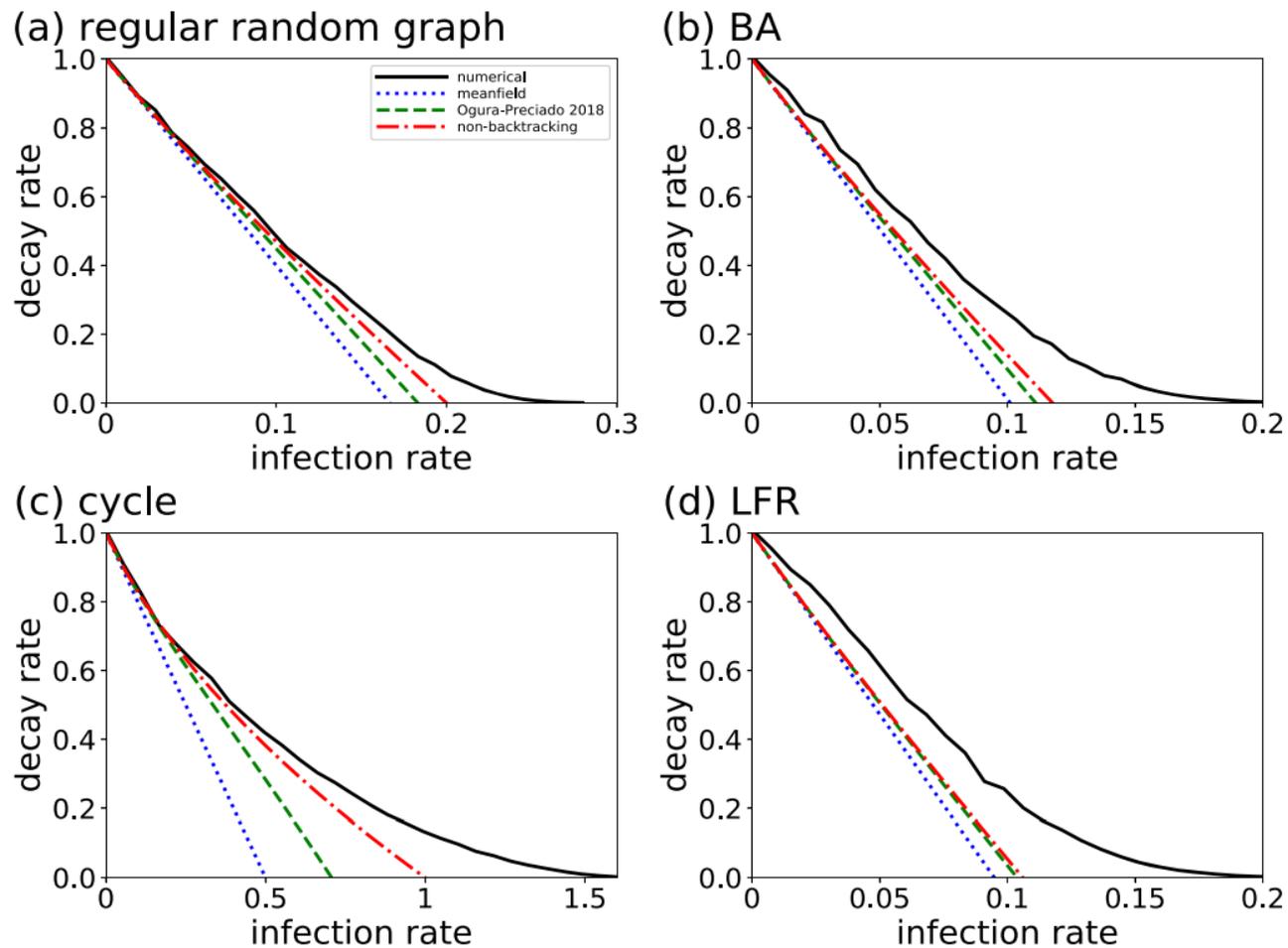
要素 1 を配置



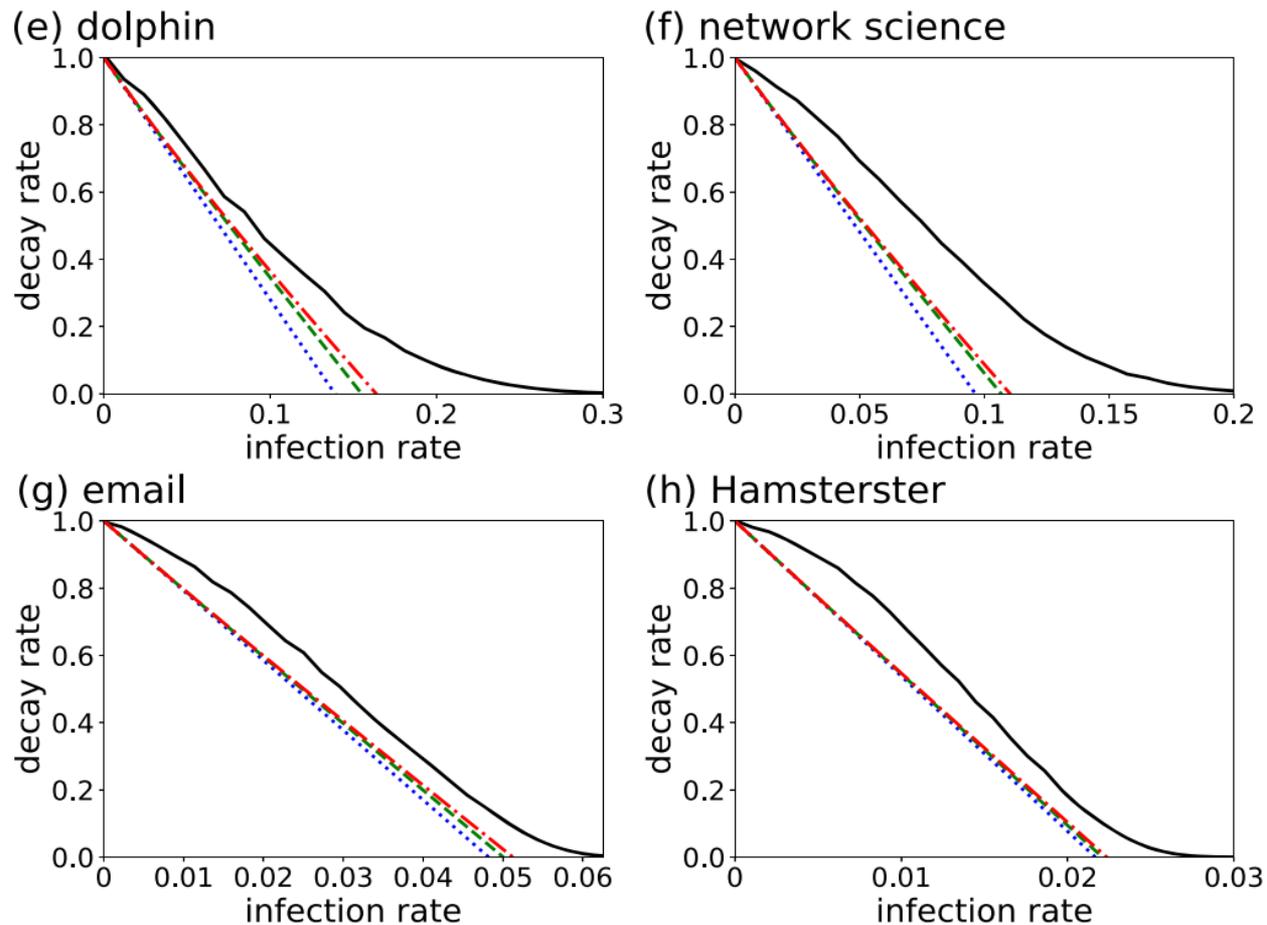
	1→2	1→3	2→1	2→3	3→1	3→2	3→4	4→3	
0	0	0	1	0	0	0	0	0	1→2
0	0	0	0	0	0	1	1	0	1→3
0	1	0	0	0	0	0	0	0	2→1
0	0	0	0	1	0	1	0	0	2→3
1	0	0	0	0	0	0	0	0	3→1
0	0	1	0	0	0	0	0	0	3→2
0	0	0	0	0	0	0	0	0	3→4
0	0	0	0	1	1	0	0	0	4→3

(b) 非バックトラック行列

Comparison of the decay rate and its lower-bound



Comparison of the decay rate and its lower-bound



中心性

モチベーション

- ネットワークは複雑でよくわからない
- どの頂点が重要か, だけでも知ることはできないか?
- 例
 - オンラインソーシャルネットワークにおける**影響力の大きいアカウント**
 - インターネットにおける**重要なウェブページ**

中心性 = 頂点のスコア

- いろいろなスコアのつけ方
 - 次数中心性: 辺の多寡により評価
 - 媒介中心性: 2つの頂点を結ぶ頻度により評価
 - 近接中心性: 他のノードに対する近さを用いて重要性を評価
 - PageRank: Google 社創業のきっかけ(?)
 - 固有ベクトル中心性: 本日のテーマ

アイデア

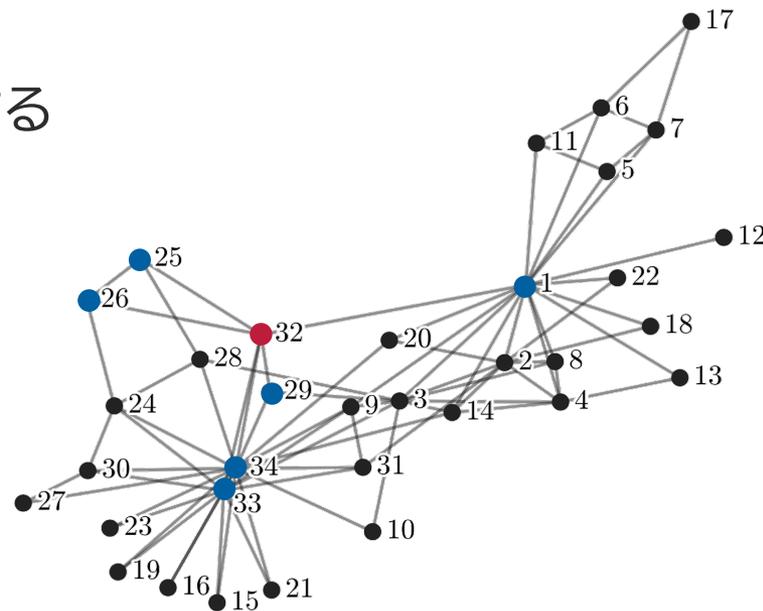
- 自分の中心性は周りの中心性の和に比例する
- 例: $x_{32} = \lambda(x_1 + x_{25} + x_{26} + x_{33} + x_{34})$
 - x_i が中心性, λ が比例定数

固有値方程式

$$\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} = \lambda A \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}$$

↑
隣接行列

- 中心性は固有ベクトルにより与えられる



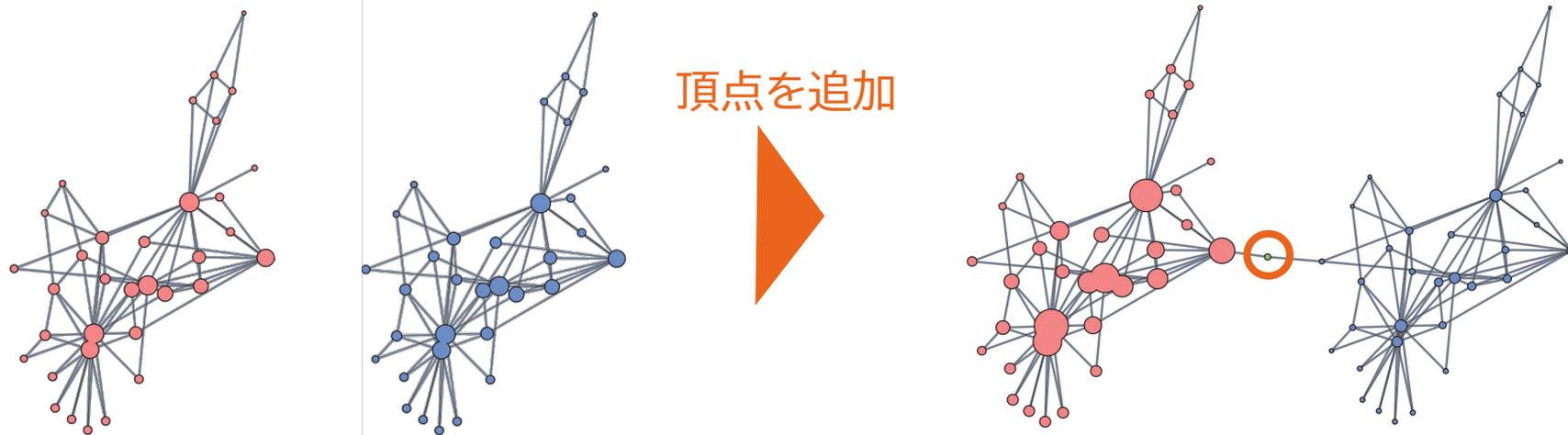
利点

- 重みが $\{0, 1\}$ や負以外の場合へも拡張可能

欠点

- 中心性の局在化

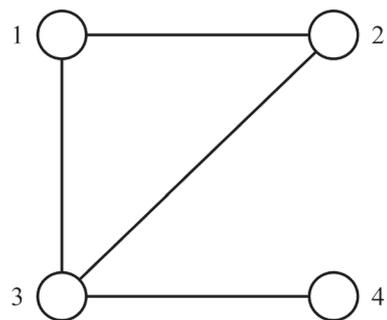
- “Richer gets richer”, 「お前のものは俺のもの」



- バックトラックが原因(と考えられる)

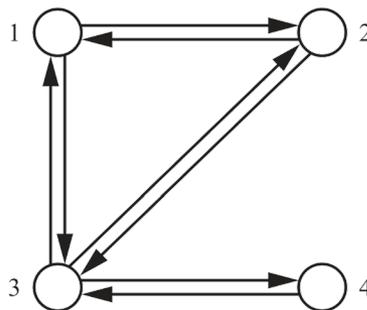
中心性が左に局在
→ なにかおかしいと考えられる

バックトラックではないパスを作る有向辺の組を関係付ける

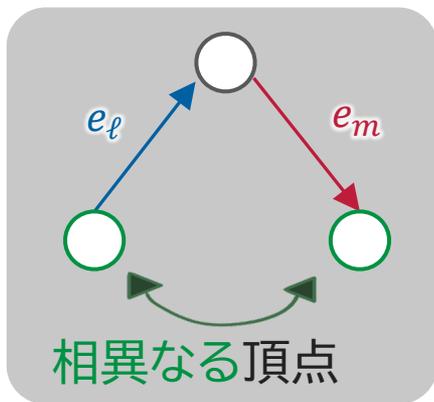


(a) 無向グラフ

辺を分解



(a) 有向グラフ D_G



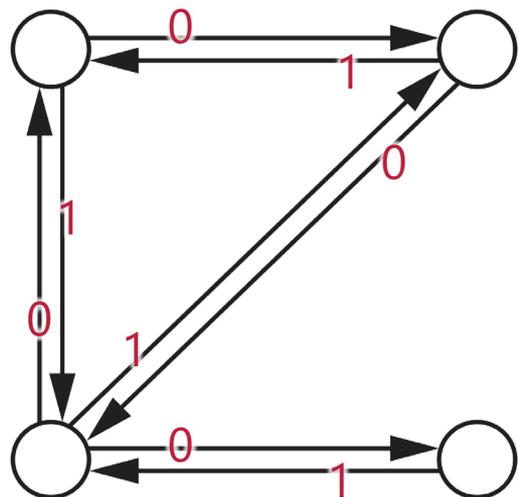
要素1を配置



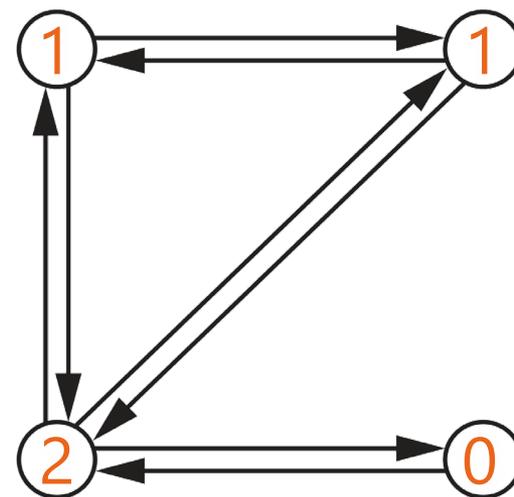
	1→2	1→3	2→1	2→3	3→1	3→2	3→4	4→3	
$B =$	0	0	0	1	0	0	0	0	1→2
	0	0	0	0	0	1	1	0	1→3
	0	1	0	0	0	0	0	0	2→1
	0	0	0	0	1	0	1	0	2→3
	1	0	0	0	0	0	0	0	3→1
	0	0	1	0	0	0	0	0	3→2
	0	0	0	0	0	0	0	0	3→4
	0	0	0	0	1	1	0	0	4→3

(b) 非バックトラック行列

非バックトラック行列の固有ベクトルを用いた中心性



値を終点へ集約

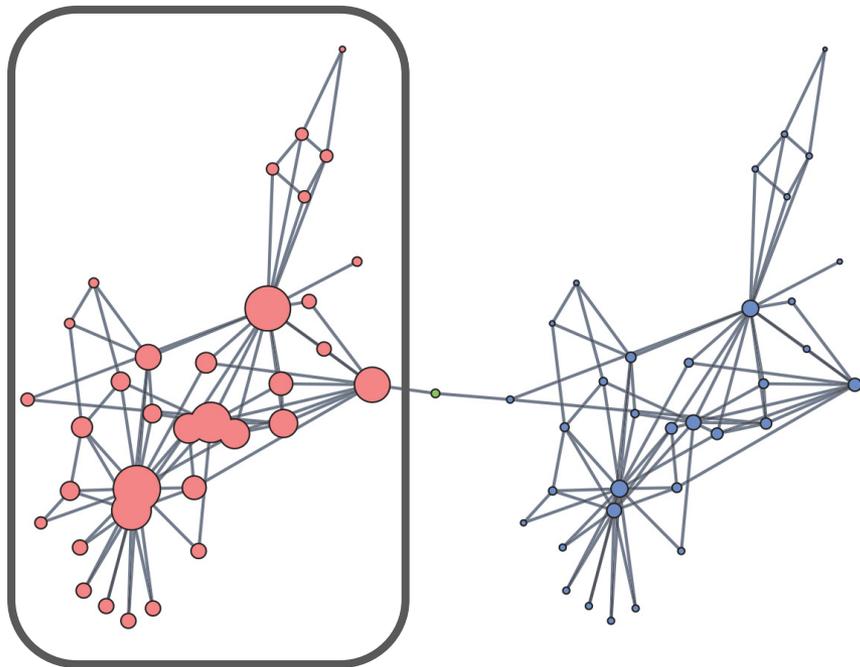


非バックトラック行列の固有ベクトル

非バックトラック固有ベクトル中心性

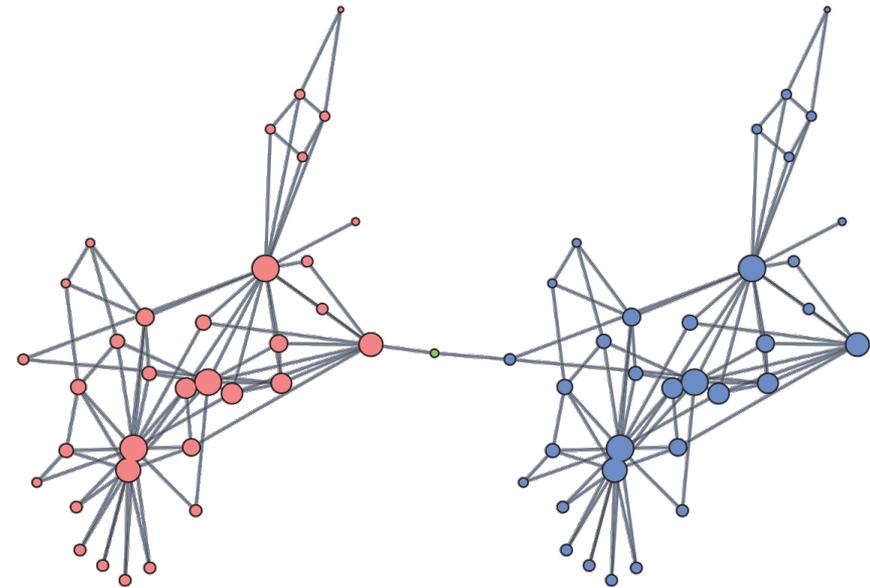
さきほどの例

固有ベクトル中心性



中心性の偏り

非バックトラック固有ベクトル中心性



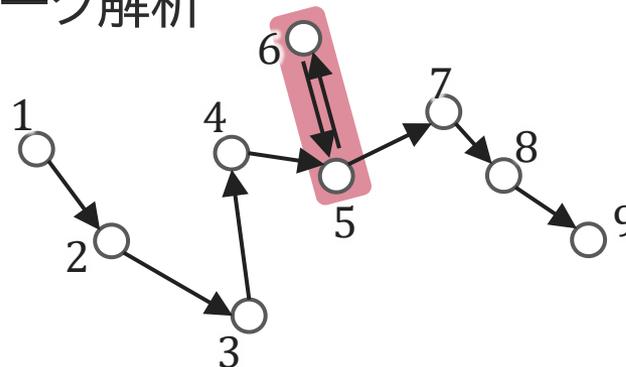
偏りを回避

おわりに

おわりに

非バックトラックに関連した複雑ネットワーク解析

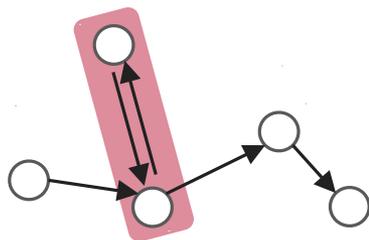
- 酔歩
- 伝播
- 中心性
- 様々な応用が期待される



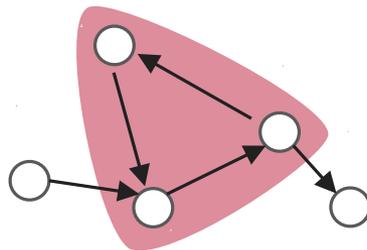
面白いトピック(の一つ)

- バックトラックの一般化

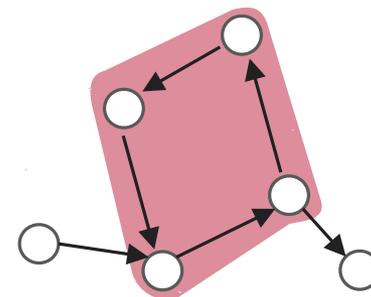
back-tracking



triangulating



squaring



小蔵, "複雑ネットワーク解析における非バックトラック," 電子情報通信学会誌, vol. 105, no. 1, pp. 27-32, 2022

Arrigo, Higham, and Voferini, "Beyond non-backtracking: non-cycling network centrality measures," Proc. R. Soc. A Math. Phys. Eng. Sci., 2020.