ネットワーク分析入門

東北大学経済学部1年 澤谷一磨

今回の内容

- 1. 基本事項の整理
- 2. ネットワークの魅力
- 3. 分析のための知識
- 4. 実装してみた(おまけ)

今回の内容

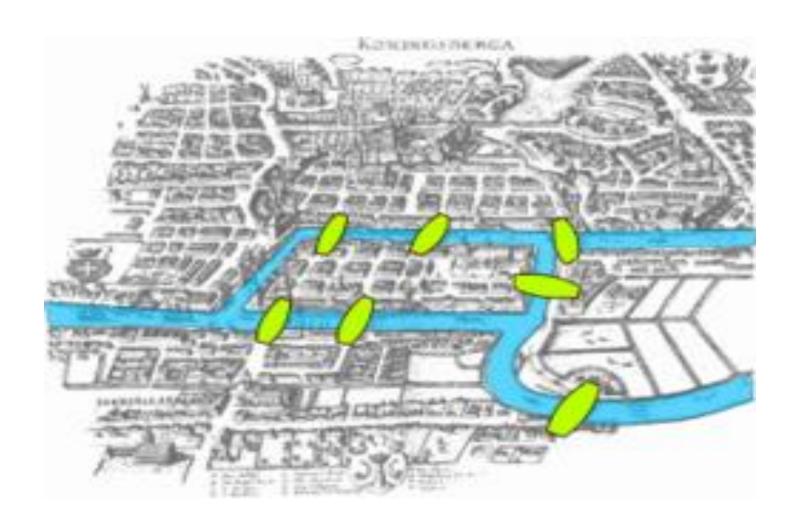
1. 基本事項の整理

2. ネットワークの魅力

3. 分析のための知識

4. 実装してみた(おまけ)

ケーニヒスベルクの橋

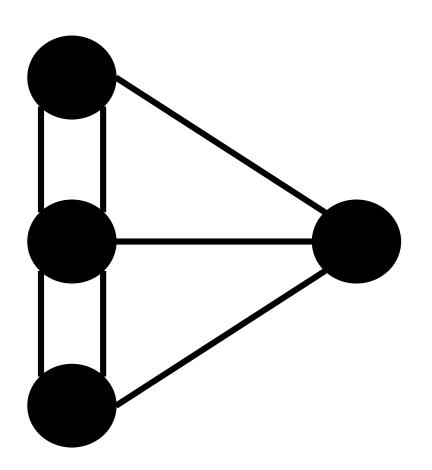


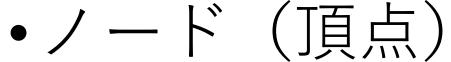
(Wikipediaより)

ケーニヒスベルクの橋

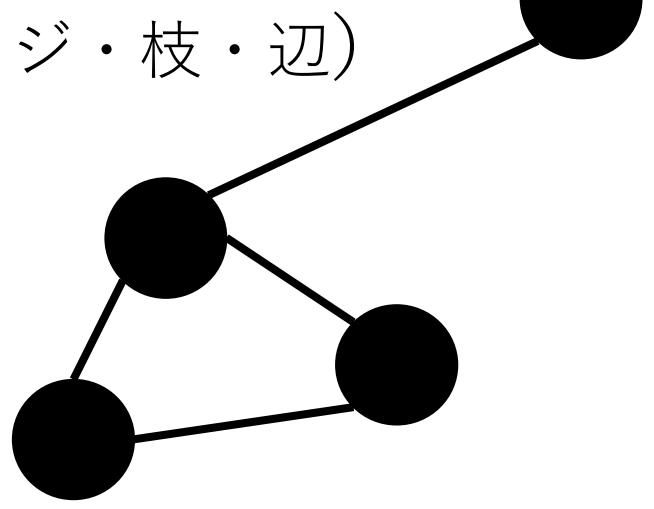


(Leonhard Euler, 1707–1783)





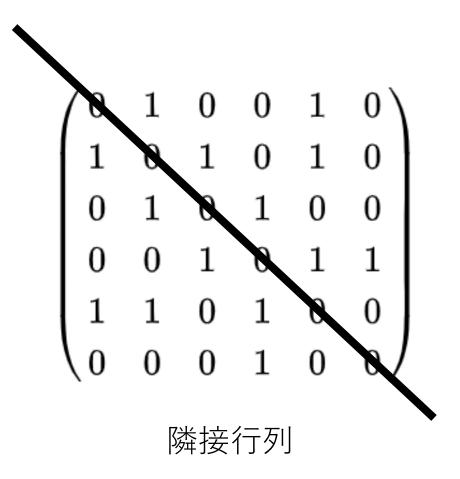
- •リンク (エッジ・枝・辺)
- •次数

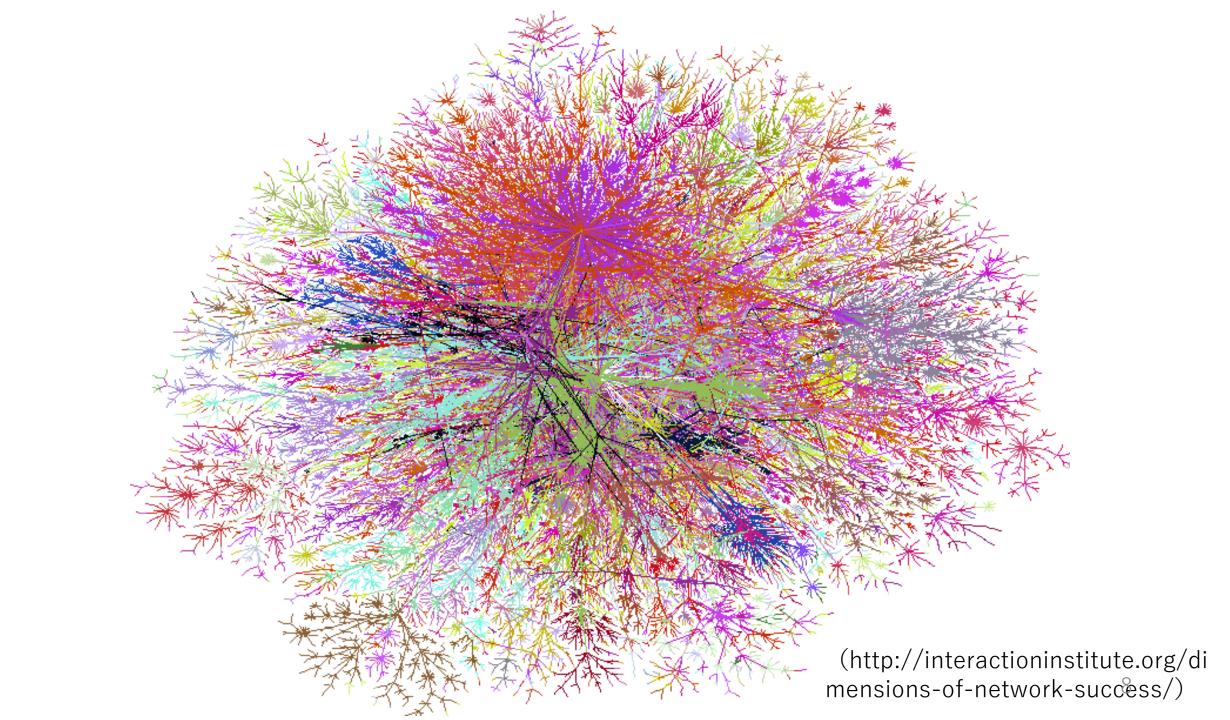


様々なグラフ

- 有向グラフ/無向グラフ
- 重み付きグラフ
- 完全グラフ
- 二部グラフ
- 木
- オイラーグラフ
- ハミルトングラフ

• • • • •





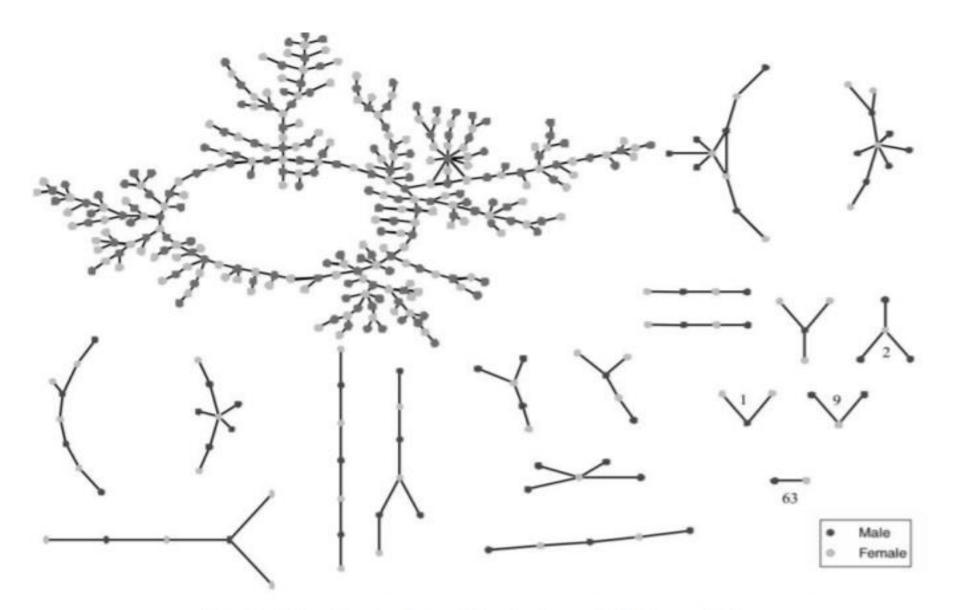
今回の内容

1. 基本事項の整理

2. ネットワークの魅力

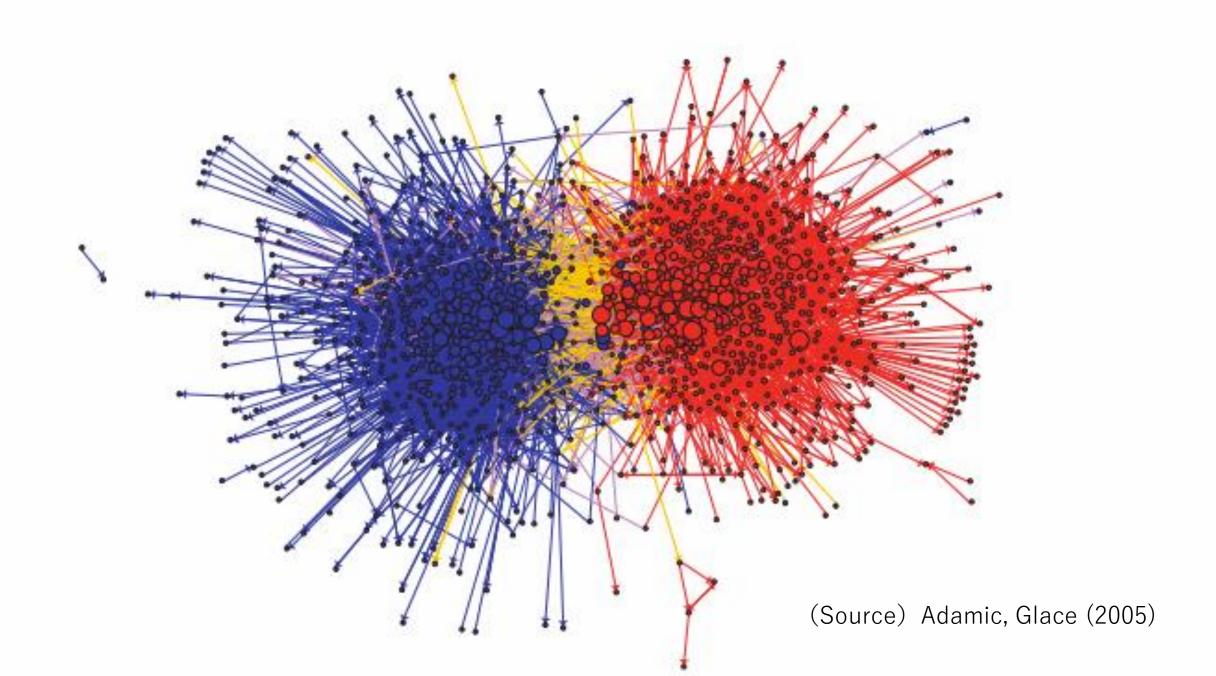
3. 分析のための知識

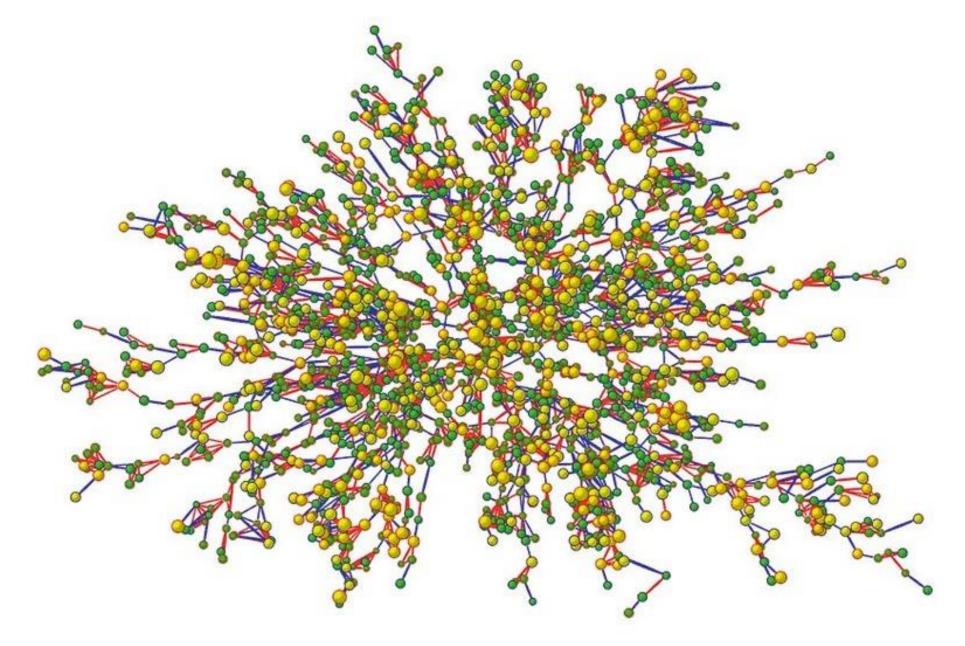
4. 実装してみた



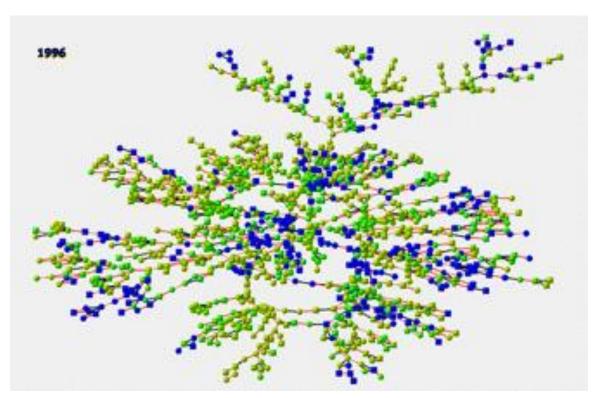
(James Moody, Katherine Strobel, Peter S. Bearman, 2004)

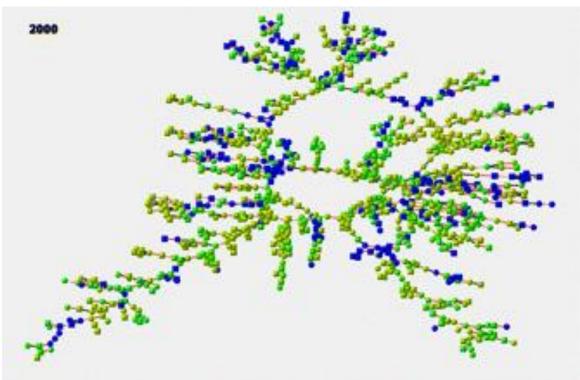
Fig. 2.—The direct relationship structure at Jefferson High





(Nicholas A. Christakis, James H. Fowler, 2007)





(Nicholas A. Christakis, James H. Fowler, 2008)

(IPSJ magazine Vol.49 No.3 Mar. 2008)

21. 日本臨床救急医学会等

17. 日本整形外科学会等

1. 電子情報通信学会等

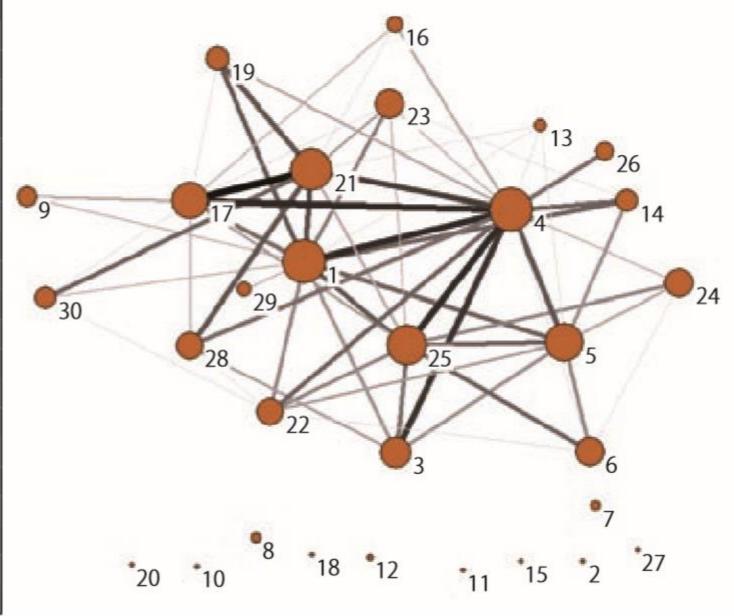
3. 日本社会情報学会等

5. 歴史人類学会等

4. 日本教育心理学会等

25. 日本教育工学会等

#	ラベル	学会数	エッジ数
1	理工	413	1,033
3	社会・福祉	72	152
4	心理・経済	365	1,016
5	史学・哲学・宗教	197	650
6	文学	59	181
9	医学(眼科)	14	43
13	映像・写真	6	5
14	農業経済	22	56
16	油脂	10	11
17	医学(整形外科・神経)・薬学	141	450
19	獣医畜産	23	69
21	医学(内科)	322	1,429
22	服飾・繊維	48	70
23	医学(生理学)・細菌	56	89
24	比較文化	51	81
25	文化・言語・教育	218	586
26	図書館	11	27
28	看護	43	142
29	工学(省略)	8	7
30	医学(細胞)	14	43



その他さまざまな研究

- 故障や攻撃に強い回線のネットワーク構築法(工学)
- 特許引用ネットワークから知財権の強化へ(法律)
- 体内の代謝ネットワーク解析(生物学)
- 仮想水のネットワークから水資源の枯渇を抑止する(環境)
- 最適なサプライチェーン構造(経営)
- ・ハリウッドの共演関係(芸能)
- 感染症に強いネットワーク構築(医学)
- ・地域通貨の流通(経済)

Etc.

ネットワークにはある共通した3つの特質があることがわかった

スモールワールド性

6次の隔たり (Duncan.J.Watts,2002)

Q.ランダムに選ばれた二人は何人を介してつながるか

その他の例(Barabasi,A-L, 2002)

- 食物連鎖(L ≅ 2)
- ・細胞中の分子の化学反応ネットワーク(L≥3)
- 科学論文の共著関係ネットワーク(L≃4~6)
- ●線虫の脳内神経ネットワーク(L ≅ 4)
- WWW (L≅19)

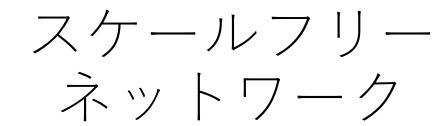
スモールワールド現象の所以

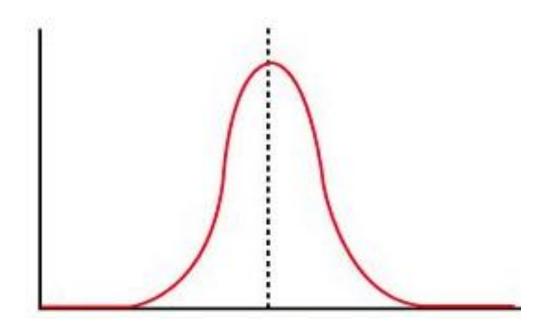
ネットワークにおける距離はユークリッド距離 と異なる

- ランダムネットワークモデルの例:
 - \bullet nステップで k^n 個の枝に到達可能
 - $L \simeq \frac{logN}{logk}$ (N:頂点数, k:次数平均)

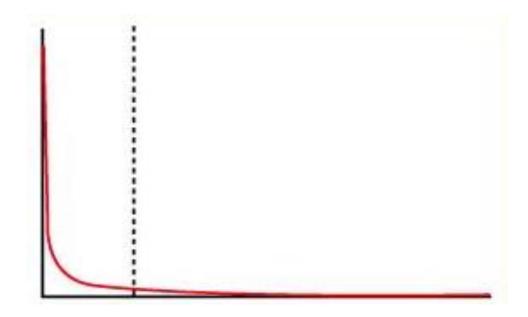
スケールフリー性

ランダムネットワーク



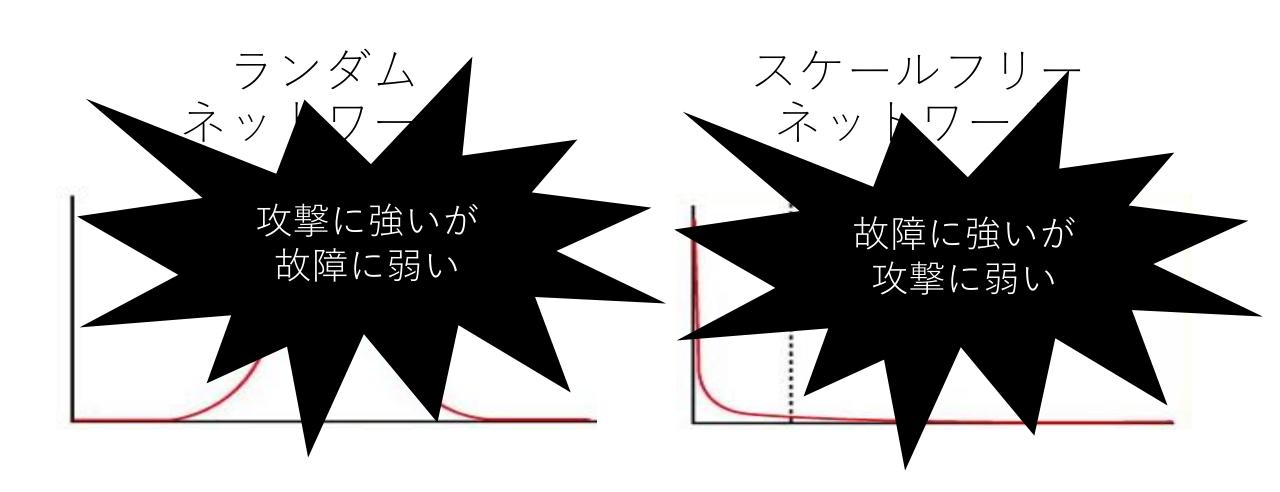


$$P[X=k]=inom{n}{k}p^k(1-p)^{n-k}\quad ext{for }k=0,1,2,\ldots,n$$

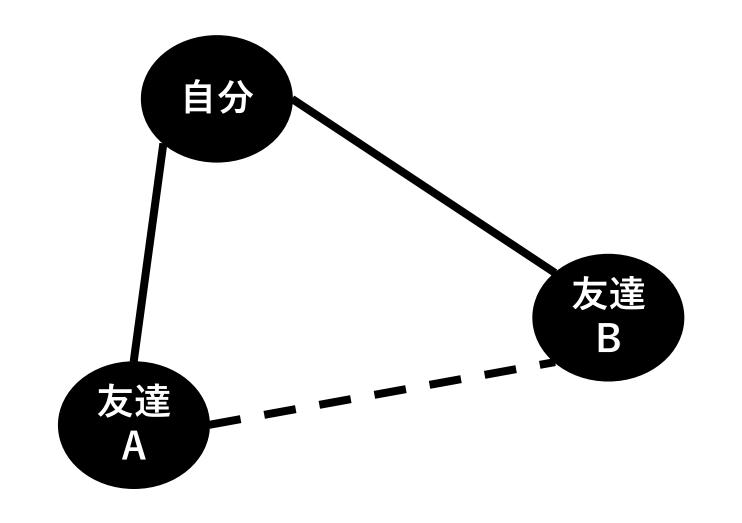


$$P(k) \propto k^{-\gamma} \ (k >> 1)$$

ネットワークの頑強性



クラスター性



 $\frac{クラスター=三角形}{(多変量解析のクラスターとは異なる)}$

今回の内容

- 1. 基本事項の整理
- 2. ネットワークの魅力
- 3. 分析のための知識
- 4. 実装してみた

ネットワークの特徴量

- 1. 次数分布·平均次数
- 2. 密度
- 3. 最短距離(幅優先探索・ダイクストラ法)・平均距離
- 4. 直径
- 5. 連結/非連結
- 6. クラスター係数 (ノードorネットワーク)
- 7. 推移性
- 8. コミュニティ構造(ギルバン=ニューマン、モジュラリティ)
- 9. 中心性
 - 次数中心性・近接中心性・媒介中心性・固有ベクトル中心性・Page Rank etc.

各特徴量の示すもの

- ネットワークの全体像
 - 直径、密度、次数分布、他
- スモールワールド性
 - 平均距離、直径
- スケールフリー性
 - 次数分布及びbin調整
- クラスター性
 - クラスター係数

- ネットワーク内の派閥形成
 - コミュニティ構造
- 重要ノードの測定
 - 中心性(次数、離心、近接、媒介、 固有ベクトル、PageRank、情報、 ボナチッチのパワー)

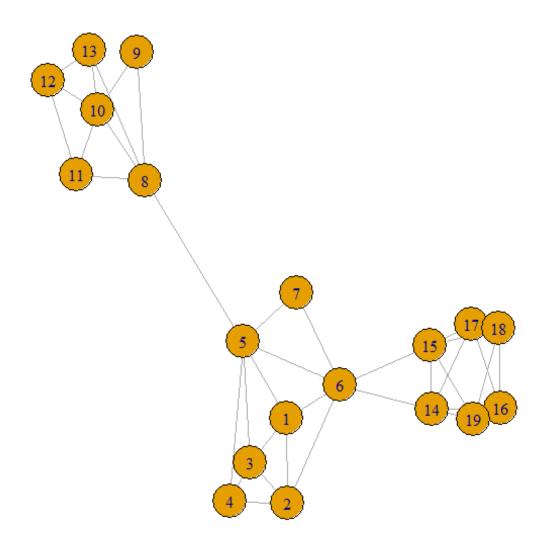
中心性――各要素の重要度を測る

- 次数中心性
 - 友達が多い奴が偉い
 - centr_degree()

- 近接中心性
 - 他の奴らとの距離が近い 奴が影響力があって偉い
 - closeness()

- 媒介中心性
 - 最短経路上にいる率が高い奴が結節点だから偉い
 - betweenness()
- 固有ベクトル中心性
 - 自分だけじゃなく偉い友 達を持ってる奴が偉い
 - centr_eigen()

コミュニティ検出――いわゆる「クラスター分析」



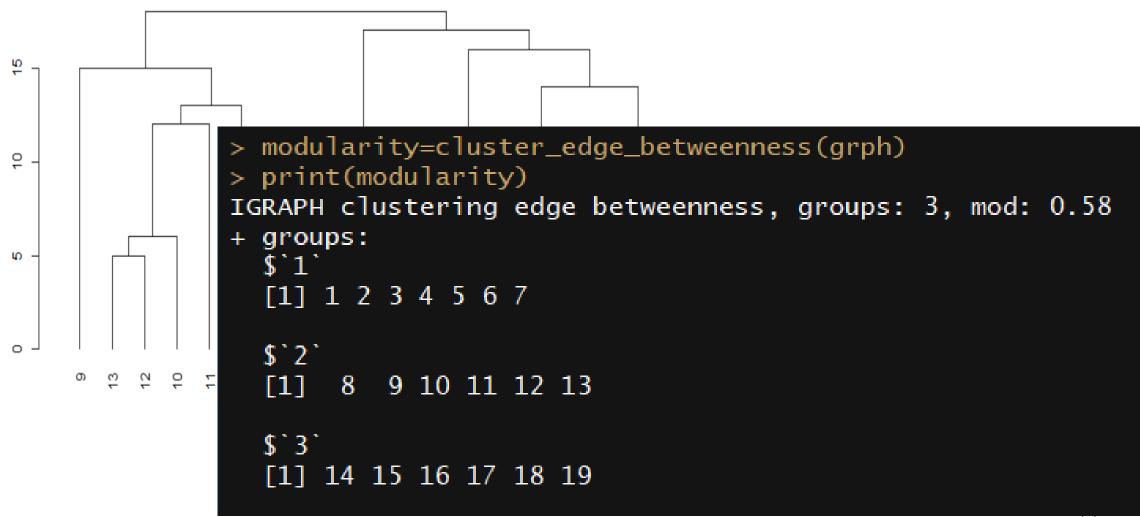
ギルバン=ニューマン法 コミュニティ数を決めておき、連結成分が その数になるまで媒介中心性最大の枝を順 に取り除いてゆく

モジュラリティを用いる方法 事前にコミュニティ数を決めず、

$$Q = \frac{1}{2M} \sum_{s=1}^{N_c} [\sum_{i,j=1}^{N} (a_{ij} - \frac{k_i k_j}{2M})]$$

においてQmaxとなるNcを定義する

コミュニティ検出



古典的なネットワークモデル

- ランダム
 - ランダム・ネットワークモデル
 - コンフィギュレーション・モデル
 - レギュラー・ランダム・ネットワークモデル
- スモールワールド
 - WSモデル
- スケールフリー
 - BAモデル

今回の内容

- 1. 基本事項の整理
- 2. ネットワークの魅力
- 3. 分析のための知識

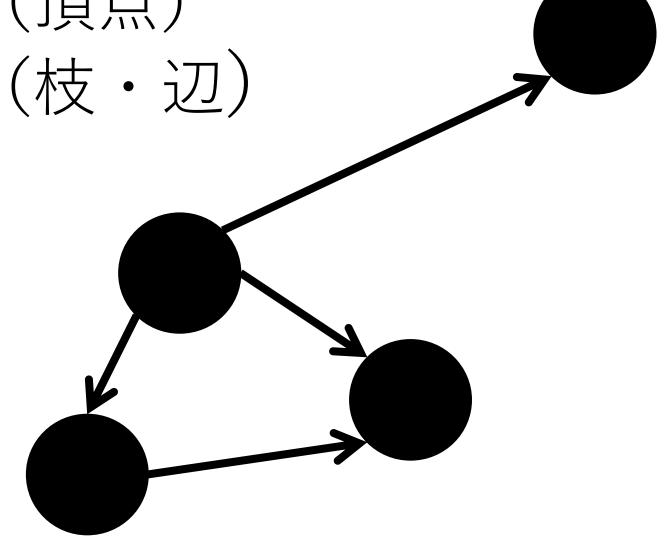
4. 実装してみた

哲学者たちのネットワーク



• リンク (枝・辺)

• 次数



1.調查目的

以下の問いの検証を目的とする

01

哲学者の影響ネットワークと一般のネットワークの特徴は符合するか否か。

02

哲学者の重要性の程度, ひいては学習上の優先度 を図れるのではないか。

対応する仮説は以下の通り。

01

部分的には当てはまるが、部分的には当 てはまらない 02

一定程度可能

||.調査手順

調査手順

1	R StudioのlibraryであるSPARQLを通じてDBpediaに格納されたLOD(Linked Open Data)を抽出し,データフレームとして保存する。
2	R Studioのlibraryであるigraphを用いてグラフの概形を知るとともに特徴量を 抽出する。
3	同データからcsv fileを作成し,Gephiを通じてよりグラフィカルに見やすい形でグラフを描画し,より深い洞察を試みる。
4	また、この際、データ収集に漏れた重要リンクを追加し、リンク数1のノードのフィルタリングやラベルの書式統一を行ってデータを修正した。

ワイキペティア フリー百科事典

メインページ コミュニティ・ポータ 最近の出来事 新しいページ 最近の更新 おまかせ表示 練習用ページ アップロード (ウィキメ ディア・コモンズ)

ヘルプ

ヘルプ 井戸端 お知らせ バグの報告 寄付 ウィキペディアに関す るお問い合わせ

ツール

リンク元 関連ページの更新状況 ファイルをアップロー 特別ページ この版への固定リンク



「ニーチェ」はこの項目へ転送されています。その他の用法については「ニーチェ (曖昧さ回 避)」をご覧ください。

フリードリヒ・ヴィルヘルム・ニーチェ(独: Friedrich Wilhelm Nietzsche, 1844年10月15 日 - 1900年8月25日) は、ドイツの哲学者、古典 文献学者。現代では実存主義の代表的な思想家の一 人として知られる。古典文献学者フリードリヒ・ヴ ィルヘルム・リッチュルに才能を見出され、哲学教 授職を希望しつつも、バーゼル大学古典文献学教授 となり、辞職した後は在野の哲して一生を過 ごした。随所にアフォリズムを 巧みな散文 的表現による試みには、文学的価値 られる。 なお、ドイツ語では、「ニーチェ」(フ [ˈfriːdrɪç] ヴィルヘルム [ˈvɪlhɛlm] [ˈniːtʃə]) のみならず「ニーツシェ」[ˈnjʲ も発音される^[1]。

目次 [非表示]

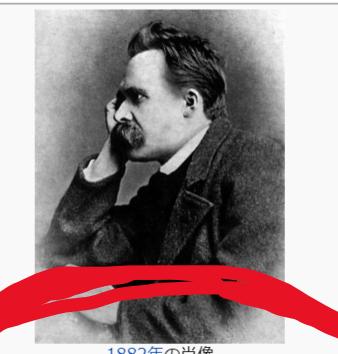
1 生涯

1.1 少年時代

1.1.1 ある雨の日の話

1.1.2 エリザベートの兄への思い

フリードリヒ・ヴィルヘルム・ニーチ Friedrich Wilhelm Nietzsche



1882年の肖像

影響を受けた人物:

[隠す]

フョードル・ドストエフスキー、ヨハン・ヴ オルフガング・フォン・ゲーテ、ラルフ・ワ ルド・エマーソン、ヘラクレイトス、ヨハ ン・ゴットフリート・ヘルダー、バーナー ド・デ・マンデヴィル. ブレーズ・パスカ





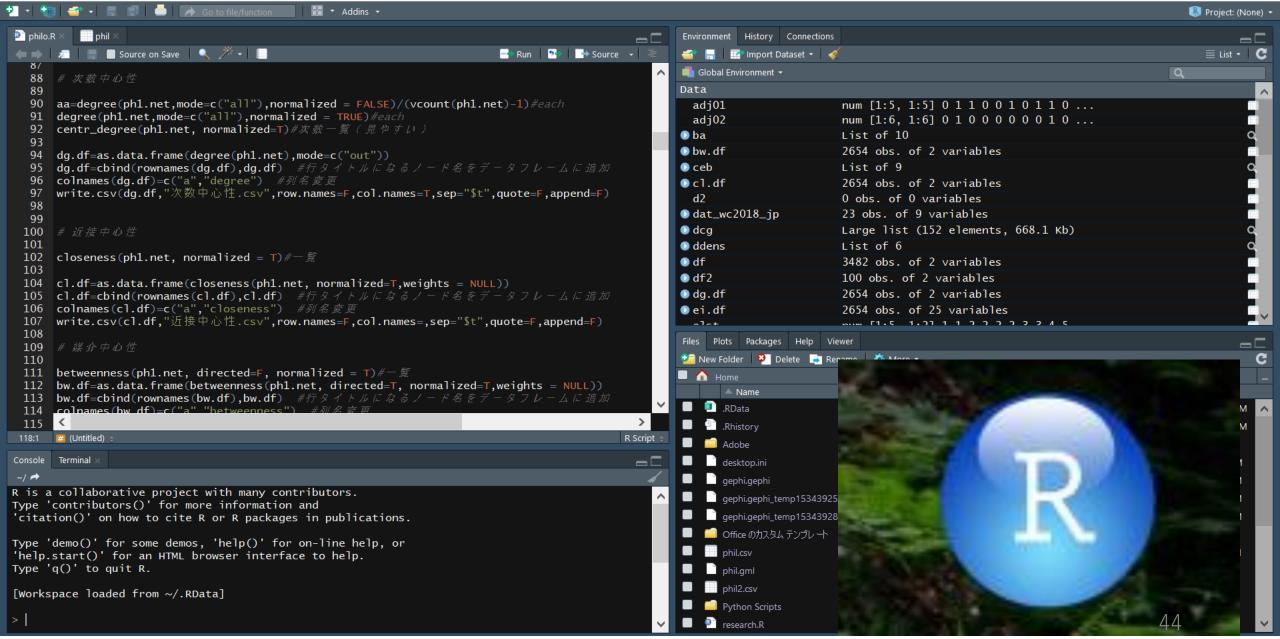


Research



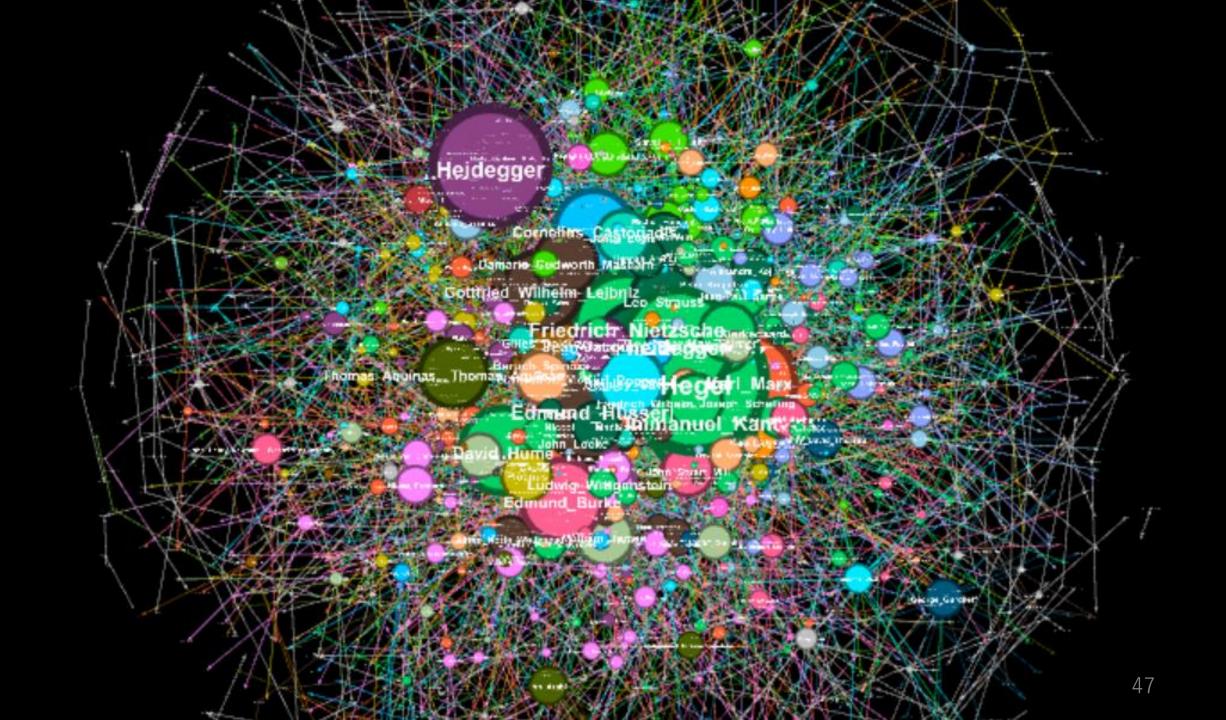
Join

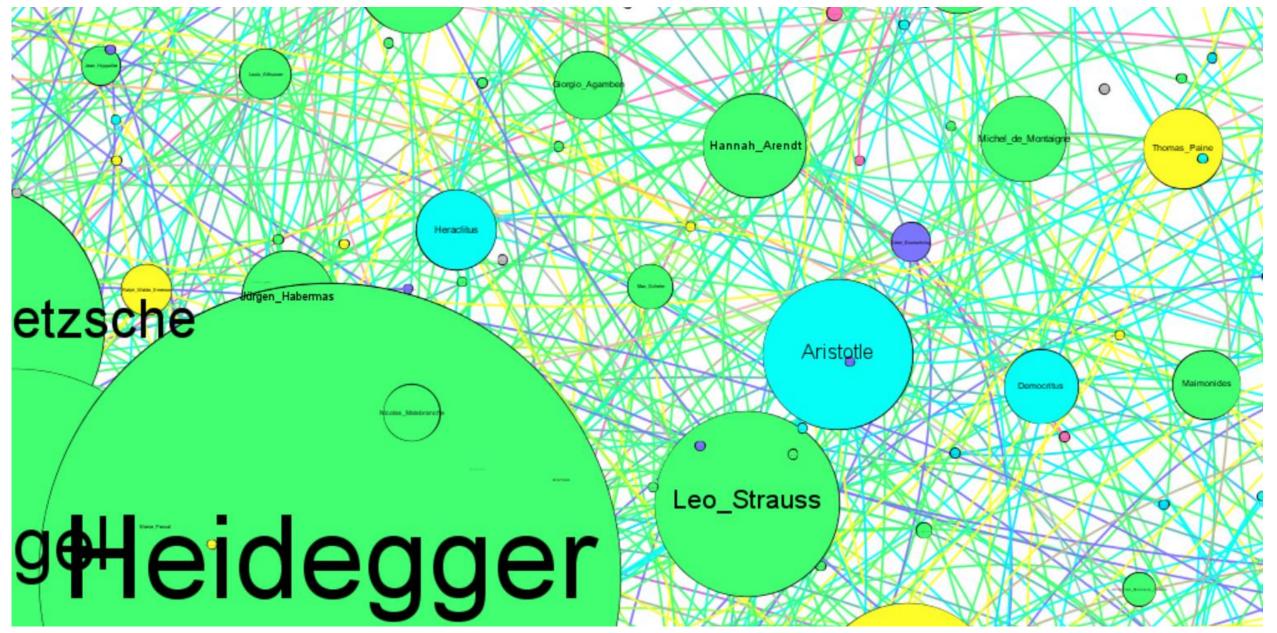
File Edit Code View Plots Session Build Debug Profile Tools Help



|||.調査結果







特徴量一覧

	R Jpn	R Eng	Gephi Jpn	Gephi Eng	Gephi En修正	regular random(R	BA model(R
ノード数			716	2577	895	1000	1000
リンク数	1095	3482	1096	3483	1902		
平均次数	3.067227	2.704466	1.531	1.352	2.619	2	2
密度	0.004301861	0.001050686	0.004	0.001	0.005	0.00205005	0.001
平均距離	186.8237	952.0165	3.793	5.51	4.369	356.1924	7.211774
直径	714	2575	9	19	11	1000	16
クラスタ係数	0.2828115	0.1211242	0.283	0.121	0.188	0.005324814	0
合計クラスタ	37	152	388	560	588	155	1
最大クラスタ	2046	2046				806	1000
最適グループ	63	200				177	39
mod	0.62	0.75				0.8	0.92

2. ネットワークモデルとの比較

密度

おおよそ大

平均距離、直径

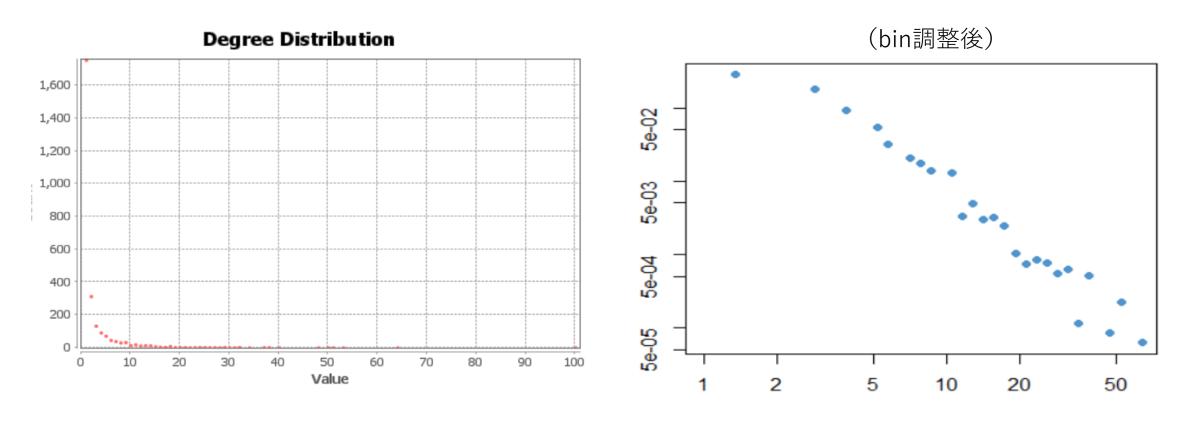
・明らかに小さい

クラスター係数

・明らかに大き

→ 現実のネットワークでは小さい平均距離と大きいクラスター係数が共存していることが多く、今回も例外ではない

3. ハブの存在



• 次数分布及びグラフより、たしかである。



4. 中心性ランキング

出次数中心性	媒介中心性	近接中心性	固有ベクトル中心性
1. ハイデガー(62)	1.ハイデガー (0.096597)	1.ハイデガー (0.001856)	1.ハイデガー(1)
2.ヘーゲル (50)	2. ヘーゲル (0.078998)	1. ヘーゲル	2.ヘーゲル (0.723582)
3. カント (36)	3.カント (0.074863)	1.カント	3.カント (0.585316)
4. カストリアデス (34)	4. ニーチェ (0.069674)	1.ニーチェ	4.ニーチェ (0.469715)
 ルソー, ライプニッツ, トマス・アクィナス, バーク(32) 	5.フッサール (0.047405)	2. ライプニッツ,フッサ ール,ルソー他21名(0.001855)	5.フッサール (0.385344)

4. 中心性ランキング

出次数中心性	媒介中心性	近接中心性	固有ベクトル中心性
1. ハイデガー(62)	1.ハイデガー (0.096597)	1.ハイデガー (0.001856)	1.ハイデガー(1)
2.ヘーゲル(50)	2. ヘーゲル (0.078998)	1. ヘーゲル	2.ヘーゲル (0.723582)
3. カント (36)	3.カント (0.074863)	1.カント	3.カント (0.585316)
4. カストリアデス (34)	4. ニーチェ (0.069674)	1.ニーチェ	4.ニーチェ (0.469715)
 ルソー、ライプニッツ、トマス・アクィナス、バーク(32) 	(0.047405)	2. ライプニッツ,フッサ ール,ルソー他21名(0.001855)	5.フッサール (0.385344)

IV.結果の考察

先行研究にみるネットワークの特徴

A 5 7 A							•			
Network	Size	$\langle k \rangle$	κ	γ_{out}	γ_{in}	l real	ℓ_{rand}	ℓ_{pow}	Reference	Nr.
WWW	325 729	4.51	900	2.45	2.1	11.2	8.32	4.77	Albert, Jeong, and Barabási 1999	1
WWW	4×10^{7}	7		2.38	2.1				Kumar et al., 1999	2
WWW	2×10^{8}	7.5	4000	2.72	2.1	16	8.85	7.61	Broder et al., 2000	3
WWW, site	260 000				1.94				Huberman and Adamic, 2000	4
Internet, domain*	3015-4389	3.42 - 3.76	30-40	2.1-2.2	2.1-2.2	4	6.3	5.2	Faloutsos, 1999	5
Internet, router*	3888	2.57	30	2.48	2.48	12.15	8.75	7.67	Faloutsos, 1999	6
Internet, router*	150 000	2.66	60	2.4	2.4	11	12.8	7.47	Govindan, 2000	7
Movie actors*	212 250	28.78	900	2.3	2.3	4.54	3.65	4.01	Barabási and Albert, 1999	8
Co-authors, SPIRES*	56 627	173	1100	1.2	1.2	4	2.12	1.95	Newman, 2001b	9
Co-authors, neuro.*	209 293	11.54	400	2.1	2.1	6	5.01	3.86	Barabási et al., 2001	10
Co-authors, math.*	70 975	3.9	120	2.5	2.5	9.5	8.2	6.53	Barabási et al., 2001	11
Sexual contacts*	2810			3.4	3.4				Liljeros et al., 2001	12
Metabolic, E. coli	778	7.4	110	2.2	2.2	3.2	3.32	2.89	Jeong et al., 2000	13
Protein, S. cerev.*	1870	2.39		2.4	2.4				Jeong, Mason, et al., 2001	14
Ythan estuary*	134	8.7	35	1.05	1.05	2.43	2.26	1.71	Montoya and Solé, 2000	14
Silwood Park*	154	4.75	27	1.13	1.13	3.4	3.23	2	Montoya and Solé, 2000	16
Citation	783 339	8.57			3				Redner, 1998	17
Phone call	53×10^{6}	3.16		2.1	2.1				Aiello et al., 2000	18
Words, co-occurrence*	460 902	70.13		2.7	2.7				Ferrer i Cancho and Solé, 2001	19
Words, synonyms*	22 311	13.48		2.8	2.8				Yook et al., 2001b	20

("Statistical mechanics of complex networks" (2002) Re ka Albert and Albert-La szlo Baraba si)

1. グラフについての考察

01

モデル・先行研究と比較しても現実のネットワークに近いネットワークが構成されていることが分かった

02

クラスター係数は小さいも のと予想され、その通り だったが想定ほど顕著なも のではなかった

3. 影響力のある哲学者ランキング

出次数中心性	媒介中心性	近接中心性	固有ベクトル中心性
1. ハイデガー(62)	1.ハイデガー (0.096597)	1.ハイデガー (0.001856)	1.ハイデガー(1)
2.ヘーゲル (50)	2. ヘーゲル (0.078998)	1. ヘーゲル	2.ヘーゲル (0.723582)
3. カント (36)	3.カント (0.074863)	1.カント	3.カント (0.585316)
4. カストリアデス (34)	4. ニーチェ (0.069674)	1.ニーチェ	4.ニーチェ (0.469715)
 5. ルソー, ライプニッツ, トマス・アクィナス, バーク(32) 	5.フッサール (0.047405)	2. ライプニッツ,フッサ ール,ルソー他21名(0.001855)	5.フッサール (0.385344)

V.まとめ



以下の問いの検証を目的とする

01

哲学者の影響ネットワークと一般のネットワークの特徴は符合するか否か。

02

哲学者の重要性の程度, ひいては学習上の優先度 を図れるのではないか。

検証の結論

01

十分に当てはまっているといえる。

02

絶対的な尺度ではあり えないものの、一定以 上の目安となりうる。

調査の改善点及び展望

データ収集の緻密化

データクレンジングの精度

万人により編集可能なWikipediaの可変性

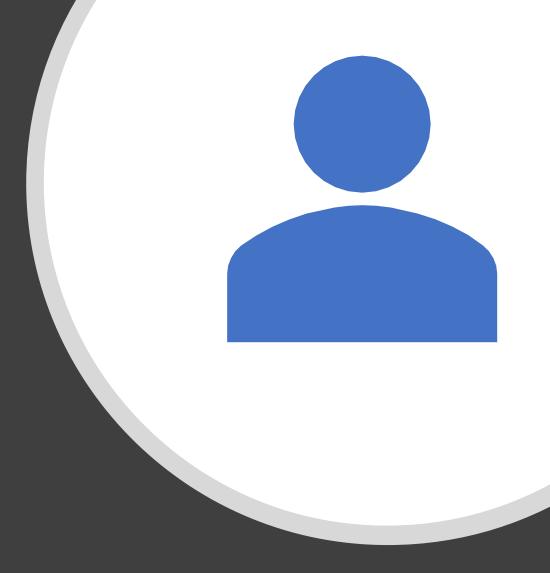
リンクの重み、すなわち影響の程度を考慮できていない

コミュニティと学派等との連関の検証

参考

文献

- "Statistical mechanics of complex networks" (2002) Re ka Albert and Albert-La szlo Baraba si
- 情報処理学会 雑誌 Vol.49 No.3 Mar. 2008 小特集「ネットワーク科学の目指すもの」
- 情報処理学会 論文誌 2017年6月号
- Rとigraphを使ったネットワーク解析と可視化 (Katherine Ognyanova 著; 田村 昌士 根本 清貴 訳)
- ネットワーク分析-Rで学ぶデータサイエンス(鈴木努、 共立出版)
- Webサイト
 - igraph R package
 - igraphの使い方
 - Gephi



その他参考になるもの

グラフ理論

- 日本の中心はどこだーグラフ理論の諸概念
 - http://www.ajimatics.com/entry/2018/01/30/133238

ネットワークサイエンス

- 「複雑ネットワークの科学」 増田直紀、今野紀雄 (産業図書)
- Network Science by Albert Barabashi
 - http://networksciencebook.com/
- Cambridge core
 - https://www.cambridge.org/core/journals/network-science

