

ネットワーク分析

首都大学東京 都市環境学部 自然・文化ツーリズムコース

倉田 陽平

ykurata@tmu.ac.jp



TOKYO METROPOLITAN UNIVERSITY

さまざまなネットワーク

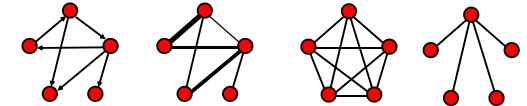


結節点のことを (またはVertex)
つながりのことを (またはEdge) という

2

ネットワークの種類

- 有向ネットワーク(directed network)
↔ 無向ネットワーク(non-directed network)
- 重みつきネットワーク(weighted network)
- 完全ネットワーク(complete network)
- ツリー(tree)



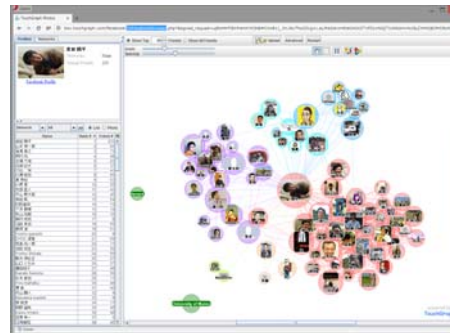
3

今日の授業の目的

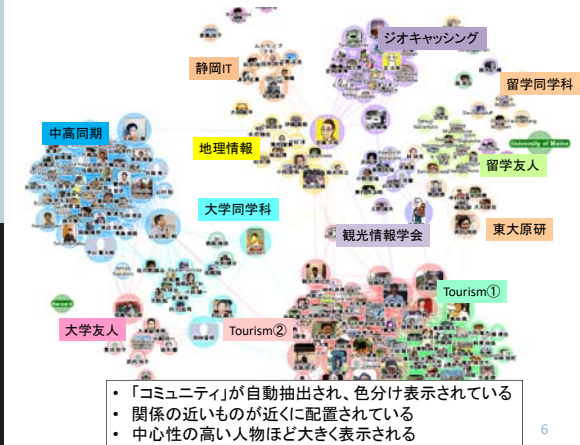
交通網・インターネット・人間関係など、社会のあちこちで登場する「ネットワーク」について基本概念と分析手法を、実際に手を動かしながら学ぶ！



倉田のFacebookの 友達ネットワークを可視化してみた ※今はできない



5



- 「コミュニティ」が自動抽出され、色分け表示されている
- 関係の近いものが近くに配置されている
- 中心性の高い人物ほど大きく表示される

6

中心性とは？

- 各ノードがネットワーク内で
どれだけ中心にあるかを示す量

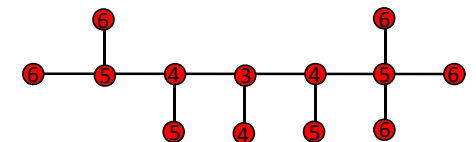


- いろいろな計算方法がある
 - 次数中心性: リンクが多いものほど中心
 - ケーニツヒ指数: 最も遠くのノードまでの距離
 - シンベル指数: ほかの全ノードまでの距離和
 - 近接中心性: (ノード数-1) ÷ シンベル指数
 - 媒介中心性: ノード間の最短経路上となる度合
 - ページランク: Googleの検索で使用

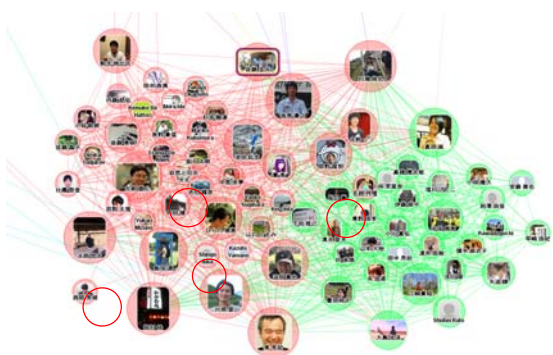
8

中心性指標①: ケーニツヒ指数

- 各ノードから最遠ノードまでのグラフ距離
- 「値が小さいノードほど中心にある」とみなす
- 最小値を「グラフの半径」という



9

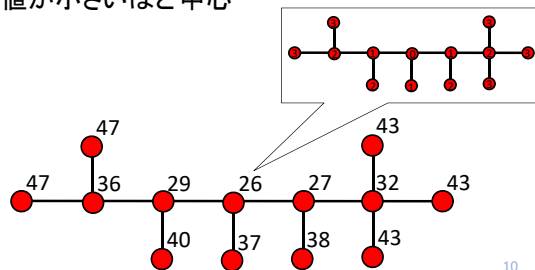


- なぜツーリズムは2グループに分かれたのか？
- なぜ先生方は中央付近に表示されているのか？
- この中にいるツーリズムでない人物は誰か？

7

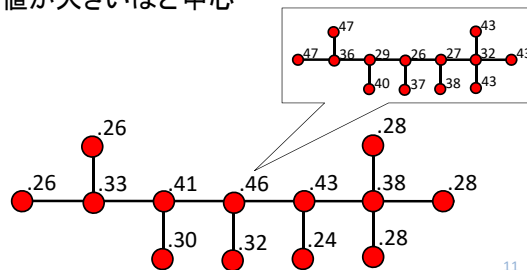
中心性指標②: シンベル指数

- 各ノードから他のノードまでのグラフ距離の和
- 値が小さいほど中心



中心性指標③: 近接中心性

- 近接中心性 = (ノード数 - 1) ÷ シンベル指数
- 値が大きいほど中心



中心性指標④: 媒介中心性

各ノードが、他のノードの組み合わせを結ぶ最短経路上に存在する割合

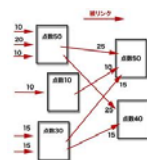
$$b_i = \frac{\sum_{l_s=1; l_s \neq i}^N \sum_{l_t=1; l_t \neq i}^{l_s-1} \frac{g_{l_s l_t}^{(i)}}{N_{l_s l_t}}}{(N-1)(N-2)/2}$$

$g_{l_s l_t}^{(i)}$: i を通る最短経路で結ばれるものの数
 $N_{l_s l_t}$: i を通らぬ最短経路の数

12

中心性指標⑤: ページランク

- Googleで使用されてきた
- 有向ネットワークが対象
- リンクを「投票」と考える
- 重要度 = 投票される量 (in) = 各ノードが投票できる量 (out)
- 「投票を集めること」あるいは「投票を集めるノードの数少ないリンク先になること」が重要度↑の秘訣



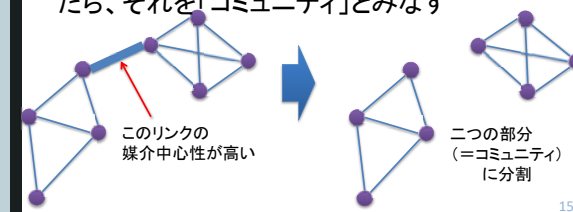
コミュニティとは

- 内部同士の結合が多く、外との結合が比較的少ない、ネットワークの部分集合 (クラスター)
- 段階構成をとる



コミュニティ算出法: ギルバン-ニューマン法

- 媒介中心性の高い (= そこが切れると多くの移動経路が遠回りになる) リンクから順に切断
- 徐々に切断をしていくうちに、独立部分が生じたら、それを「コミュニティ」とみなす



ネットワーク分析の観光への応用例

- 交通網分析
 - 地域の交通網の要衝はどこか?
- 組織分析 (まちづくり)
 - 中心人物は誰か?
 - どのようなコミュニティが内在しているか?



16

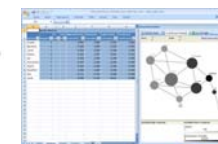
多摩動物公園をネットワーク分析してみよう



使用するツール: NodeXL

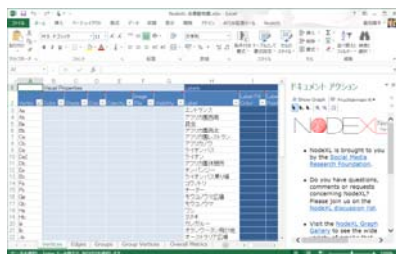
- Excel上でネットワーク分析を行うためのアドイン Basic版 (無料) と Pro版 (年額制) がある
- スタートメニューの中の「NodeXL Excel Template」もしくはNodeXLで作成したExcelファイルを直接、読み込むことで起動

Network Graphs
 The Social Media Research Foundation
<http://nodexl.codeplex.com>



多摩動物公園のデータを読む

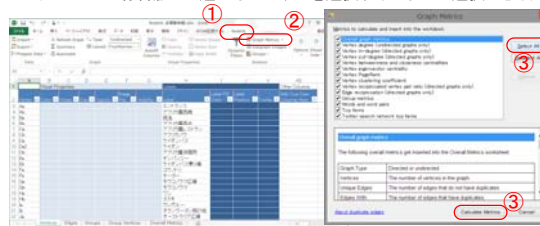
- 「17観光地理情報学特論 I データ」内にある「NodeXL-多摩動物園.xlsx」をダブルクリックして開く
- もしエラーが表示されたら、NodeXLをインストールすべし



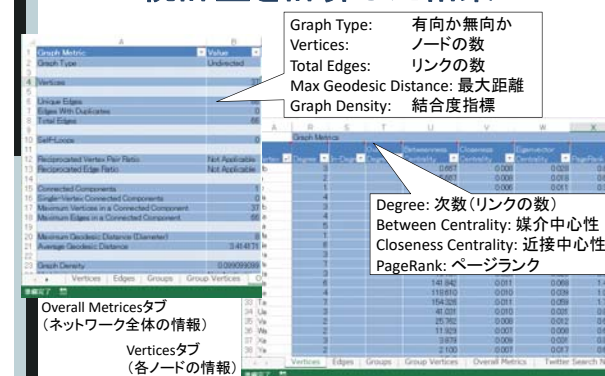
19

統計量を算出する

1. NodeXLリボンを選択
2. Graph Metricsを選択
3. Select Allを選択肢、Calculate Metricsをクリック
(いくつかは有料版しか選べないが、Allを選択するとなぜか選択できる)

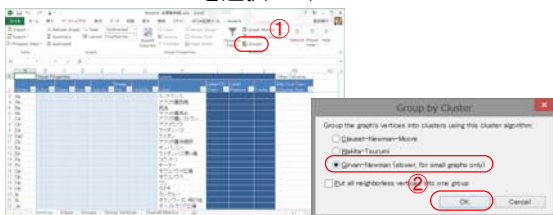


統計量を計算した結果



コミュニティを計算する

1. NodeXLリボンの中のGroups→Group by Clustersを選択
2. Girvan-Newmanを選択して、OK



22

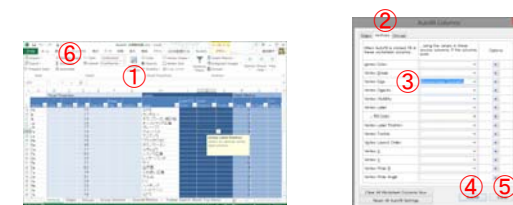
コミュニティを計算する(結果)



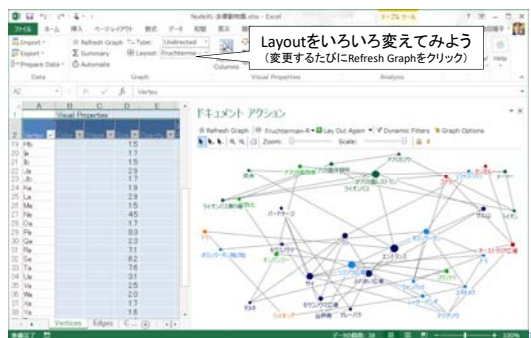
23

ネットワークの図示

- ①Autofill Columns→②Verticesタブ→③Vertex sizeをPageRankとする→④Autofill→⑤Close
- ⑥Refresh Graphボタンを押す

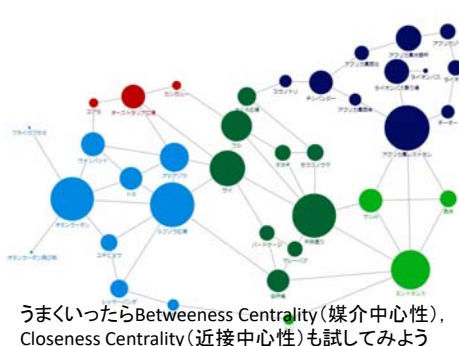


ネットワークの図示(結果)



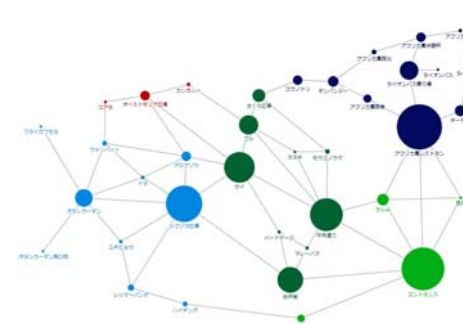
25

Page Rankによる図示



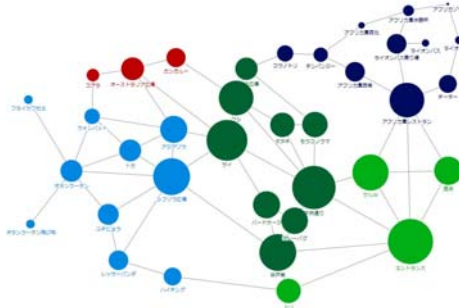
26

Betweenness Centrality (媒介中心性)による図示



27

Closeness Centrality (近接中心性)による図示



28



29

ネットワーク分析に必要なデータ

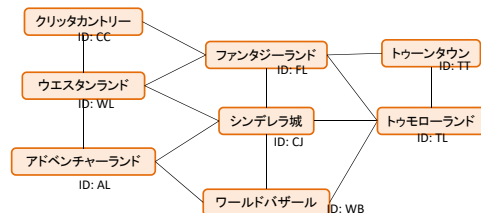
- ノードのリスト
 - ID
 - 名称
 - xy座標(オプション): 地理的データの場合
- リンクのリスト
 - ノードIDとノードIDのペア
 - ウェイト(オプション): 交通量、通信量など

30

練習:ディズニーランドを ネットワーク分析してみよう

NodeXLを使って、次のネットワークのデータを自分で入力し、分析してみよう。

どのエリアがディズニーランドの中心だろうか？



31

注意！

今回は地理的なネットワークを例に使ったが、別に実空間と関係ないネットワークでもOK

- 例: Facebookの友達関係
- HPのリンクの関係
- 同僚間のメールのやりとりの頻度
- 文章における単語の共起関係

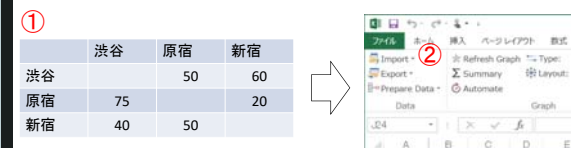
– ぶっちゃけ実空間に
関係ない方が面白い



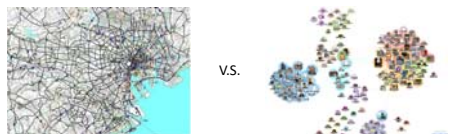
OD表のインポート

(清水研向け 参考)

- OD表の入ったExcelファイルをあらかじめ開いておく
- Import→From Open Matrix Workbook
- でてきたウィンドウで、OD表は有向(Directed)か無向(Undirected)か聞かれるので、しかるべき選択
- Importをクリック



道路ネットワークは他のネットワーク と何が違うのか？



34

スモールワールドネットワーク

- 任意の2つのノードが、中間にわずかな数のノードを介するだけで接続されるというような性質のネットワーク
- 代表例: 人のネットワーク
 - 「6次の隔たり」

- 道路網は？
- 航空網は？



スケールフリーネットワーク

- 大規模なハブ(大量のノードと繋がっているようなノード)が存在するようなネットワーク
- 代表例: インターネット

- 道路網は？
- 航空網は？



<http://internet-map.net/> 36