システム工学基礎（白山分） 2016.04-07

システムモデリングの基礎

東京大学大学院工学系研究科システム創成学専攻
東京大学システム創成学科
知能社会システムコース
白山　晋（しらやま　すすむ）
sirayama@sys.t.u-tokyo.ac.jp

講義間の関係　（システム工学基礎を中心としたとき）

数学：確率・統計，線形代数，離散数学，微分（方程式）
数理手法I　数理計画と最適化
プログラミング基礎，プログラミング応用

システム創成学基礎
データ指向モデリング
システム工学基礎
設計学基礎
システム制御工学
信頼性工学
知識と知能

社会システム工学基礎・前半
社会システム工学基礎・後半
社会システムと産業
マルチエージェントシステム　社会システム工学応用
先進デザイン　技術プロジェクトマネジメント
リスクプロジェクトマネジメント
産業組織論

システムモデリング

モデリングとは
モデルを作る（創る）こと．
モデル化ということも多い．

モデルとは
木村英紀：モデルとは何か，数理科学，9/1998，pp.5-10
・モデルとは「模型」のことである．
・モデルは実物のまねであるが，同時に実物はモデルをまねて作られる，という二面性をもっている．
・現象を定量的に解析するために導かれた数式は現象のモデルとよばれる．

	客観性	主観
理論	◎	×
モデル	○	○
芸術作品	△	◎

	普遍性	特殊性
理論	◎	△
モデル	○	○
芸術作品	△	○

モデリングの方法（留意事項）

実世界が複雑である以上，膨大なモデルが存在し（存在しない可能性もある），モデリングの方法も多種多様であると考えられる。実際これは正しいのだが，モデルはいくつかのものに分けることができる．
つまり，いくつかのモデリングの作法（定石）が存在する．
ただし，それらの分類法は一通りではないので注意されたい
（切り口の違いについても着目してほしい）。

モデリングの方法（分類）

物理現象あるいは自然現象を扱うときに用いることが多いが，システム全般における様々な現象・事象を扱うときにも使う．ただし，注目している現象や事象がどのようなモデルで扱われるか，それ自体の重要度は低い．

一次モデル（基礎方程式系モデル），
二次モデル（近似モデル），
動的モデル（非定常モデル），
静的モデル（定常モデル），
連続系モデル，離散系モデル
決定的モデル，確率的モデル
統計的モデル，
ミクロ的モデル,マクロ的モデルなど

モデリングの具体例

・要素の抽出，あるいは要素の仮定
・要素の変数化
・説明変数か被説明変数か
（独立変数か従属変数か），ただし，説明変数＝独立変数ではない
－　被説明変数は従属変数
－　物理系では，基本となる独立変数は時間と空間
－　時間と空間が説明変数となることは多くない．
－　したがって，説明変数もほとんどの場合，従属変数である．
よって，ある説明変数pは$p=p(t,x,y,z)$で表され，
ある被説明変数Qは$Q=Q(p,\cdots)$のような形で表される．
・要素（変数）を用いて何かしらの方程式を導く．あるいは方程式を選択する
方程式は等号式，あるいは不等号式
（支配方程式，あるいは制約条件）
・最適化問題へ置き換えられることも多い．

モデリングにおいて重要なこと（1）

○要素に分解できるか（要素還元の立場をとれるか）
○どこまでモデル化するか
－　主要因子を予想してモデリング
－　主要因子を調べるためのモデリング
○既存の知見，定石などを利用できるか
－　物理法則の利用
－　数学的な道具の利用（確率・統計処理など）
○要素間の相互作用が線形か非線形か，それとも予想のつかないものか
◇ヒューマンファクターが存在するか
そしてそれがモデル化できるか

モデリングにおいて重要なこと（2）

○背景に何かしらの構造が存在するか
－決定論的か確率論的か
○感度解析は可能か
○唯一のモデルか
－　複数のモデルを考え，それぞれを比較検討することは重要

■予測-仮説 →　検証
■論証の可能性，実証の必要性
■不測の事態の処理

- 相互関係のモデリング
  グラフ理論とネットワーク分析から
- 確率過程
- マルコフ連鎖（過程）
- 待ち行列

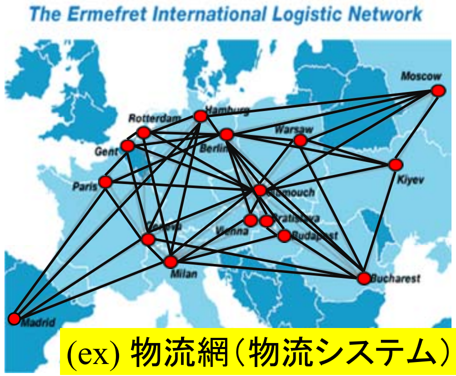
相互関係のモデリング

- グラフ理論，
- ネットワーク分析の基礎を中心に

▶ ネットワークを知る，使う，見る，作る・デザインする

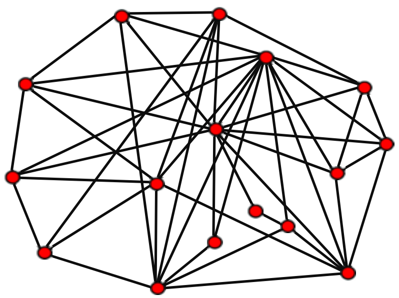
▶ ネットワーク思考

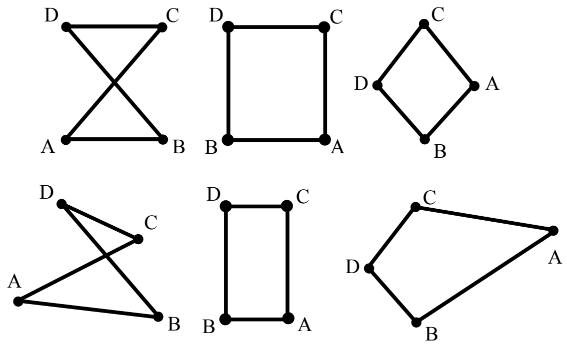
ノードに位置座標が付与されていても位相構造として扱うことが多い



(ex) 物流網（物流システム）

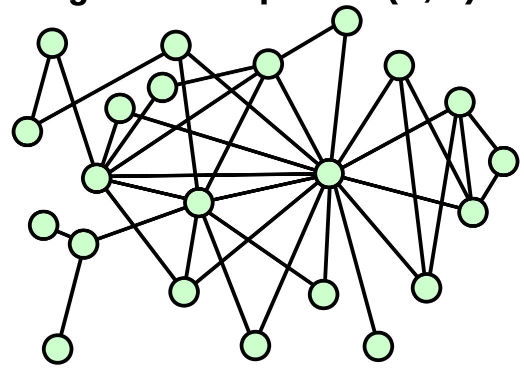
ノードに位置座標が付与されていても位相構造として扱うことが多い





位相構造のみでは，これらはすべて同じものを示す．
（物理空間の距離とグラフ距離の対応付けは難しい）

Vertices Nodes Edges > Graph G = (V,E)



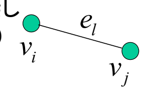
グラフ(1)

頂点：Vertex，ノード(node)
辺：エッジ(Edge)，リンク(link)，アーク(Arc)

グラフ　$G = G(V,E)$
頂点集合　V：要素は頂点　$V = V(G)$
辺集合　E：要素は辺　$E = E(G)$

$|V|$：頂点の数，位数(order)　n
$|E|$：辺の数，サイズ(size)　m

向き無し（無向）

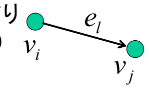


$e_l = \{v_i, v_j\}$

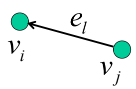
辺e_lは頂点v_iとv_jと接続している(incident)

頂点v_iとv_jは隣接している(adjacent)

向き有り（有向）



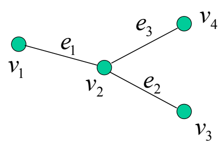
$e_l = \{v_i, v_j\}$
始点，終点



$e_l = \{v_j, v_i\}$

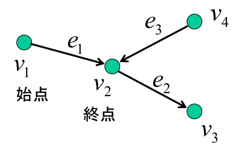
グラフ(2)

無向グラフ(undirected graph)



$V(G) = \{v_1, v_2, v_3, v_4\}$
$E(G) = \{e_1, e_2, e_3\}$
$= \{\{v_1, v_2\}, \{v_2, v_3\}, \{v_2, v_4\}\}$

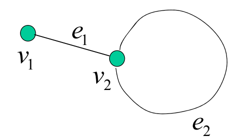
有向グラフ(directed graph)



$V(G) = \{v_1, v_2, v_3, v_4\}$
$E(G) = \{e_1, e_2, e_3\}$
$= \{\{v_1, v_2\}, \{v_2, v_3\}, \{v_4, v_2\}\}$
始点，終点

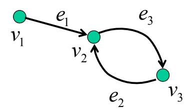
グラフ(3)

ループ(loop)　：　無向，有向が考えられる．下記は無向の例



$V(G) = \{v_1, v_2\}$
$E(G) = \{e_1, e_2\}$
$= \{\{v_1, v_2\}, \underline{\{v_2, v_2\}}\}$

多重辺（multiple edge）：　無向，有向が考えられる．

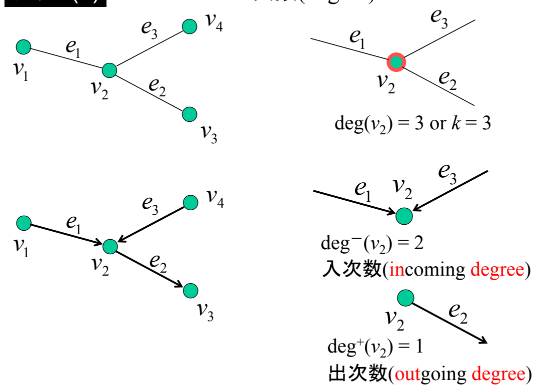


$V(G) = \{v_1, v_2, v_3\}$
$E(G) = \{e_1, e_2, e_3\}$
$= \{\{v_1, v_2\}, \underline{\{v_3, v_2\}, \{v_2, v_3\}}\}$

単純グラフ（Simple graph）：ループと多重辺を持たないグラフ

グラフ(4)

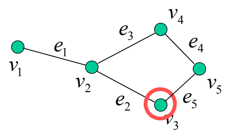
次数(degree)



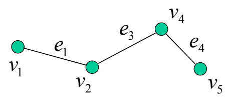
G中における次数の最大値をグラフの最大次数

グラフ(5)

頂点の除去



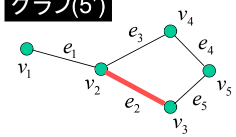
$V(G) = \{v_1, v_2, v_3, v_4, v_5\}$
$E(G) = \{e_1, e_2, e_3, e_4, e_5\}$
$= \{\{v_1, v_2\}, \{v_2, v_3\}, \{v_2, v_4\}, \{v_4, v_5\}, \{v_3, v_5\}\}$



$V(G) = \{v_1, v_2, v_4, v_5\}$
$E(G) = \{e_1, e_3, e_4\}$
$= \{\{v_1, v_2\}, \{v_2, v_4\}, \{v_4, v_5\}\}$

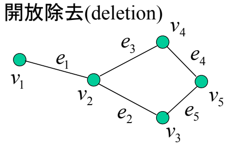
グラフ(5')

辺の除去



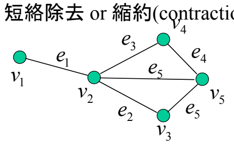
$V(G) = \{v_1, v_2, v_3, v_4, v_5\}$
$E(G) = \{e_1, e_2, e_3, e_4, e_5\}$
$= \{\{v_1, v_2\}, \{v_2, v_3\}, \{v_2, v_4\}, \{v_4, v_5\}, \{v_3, v_5\}\}$

開放除去(deletion)



$V(G) = \{v_1, v_2, v_3, v_4, v_5\}$
$E(G) = \{e_1, e_3, e_4, e_5\}$
$= \{\{v_1, v_2\}, \{v_2, v_4\}, \{v_4, v_5\}, \{v_3, v_5\}\}$

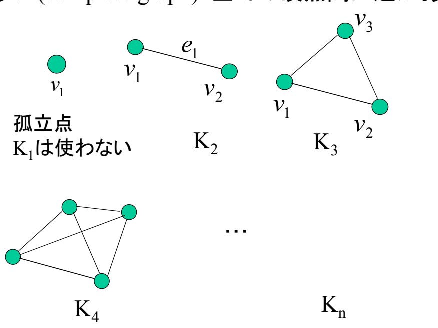
短絡除去 or 縮約(contraction)



$V(G) = \{v_1, v_2, v_3, v_4, v_5\}$
$E(G) = \{e_1, e_3, e_4, e_5\}$
$= \{\{v_1, v_2\}, \{v_2, v_4\}, \{v_4, v_5\}, \{v_2, v_5\}\}$

グラフ(6)

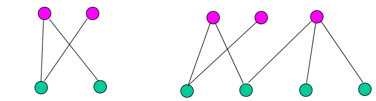
完全グラフ (complete graph)：全ての頂点間に辺があるグラフ



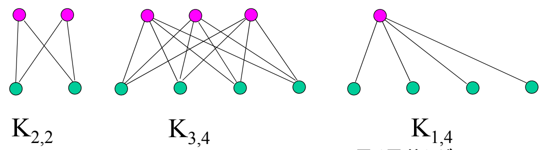
クリーク

グラフ(7)

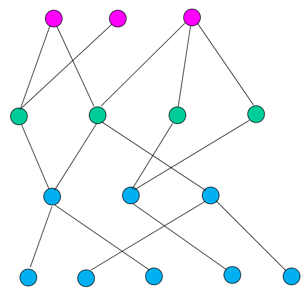
2部グラフ (bipartite graph)



完全2部グラフ (complete bipartite graph)

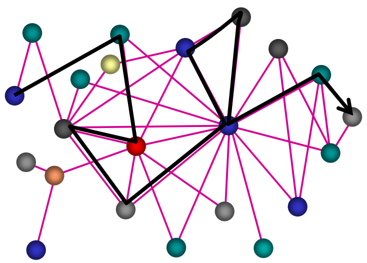


N部グラフ



グラフ(8)

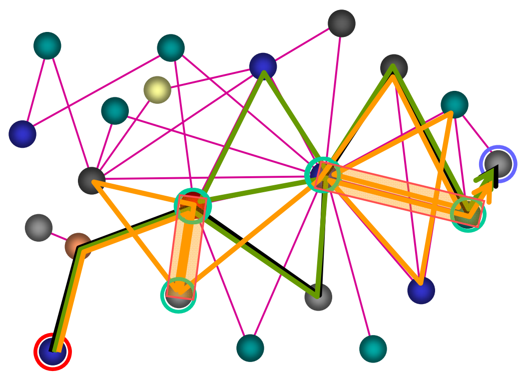
経路，道



路（パス(path)），道：初等的な道
同じ辺，同じ頂点を2度以上通らない歩道
遊歩道（トレイル(trail)），小道：単純な道
同じ辺を2度以上通らない歩道
歩道（ウォーク(walk)），歩道：道

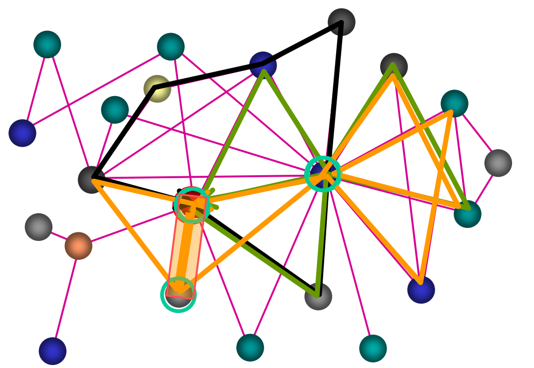
グラフ(8')

路（パス(path)），道：初等的な道
遊歩道（トレイル(trail)），小道：単純な道
歩道（ウォーク(walk)），歩道：道



グラフ(9)

閉路（サイクル(cycle)，サーキット(circuit)）：パスがベース
巡回路（ツアー(tour)）：トレイルがベース
閉ウォーク：ウォークがベース，これをサーキットと呼ぶ教科書もある



グラフ(10)

部分グラフ(subgraph)

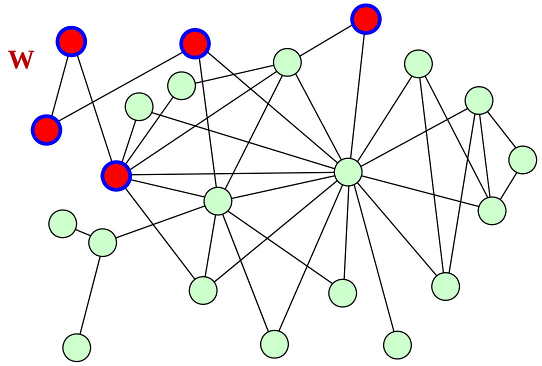
G=(**V**,**E**)のとき，**V**の空でない部分集合**W**に対して **P**$_2$(**W**)を考える．**P**$_2$(**W**)は**W**から任意の2つの要素を選んだときの（辺の）集合

E∩**P**$_2$(**W**)の部分集合を**E'**とするとき，**H**=(**W**,**E'**)を「部分グラフ」という．

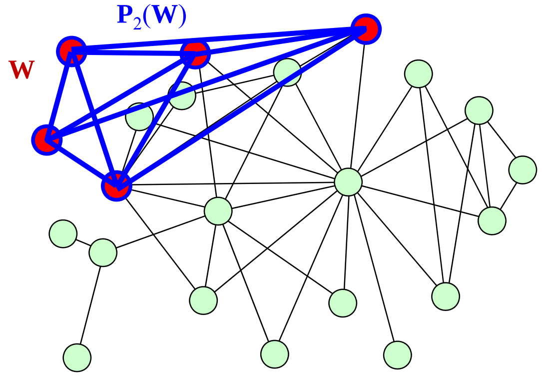
また，**G**(**W**)=(**W**, **E**∩**P**$_2$(**W**))を「誘導部分グラフ」という．

また，**V**=**W**のとき，「全域部分グラフ」という．

部分グラフ

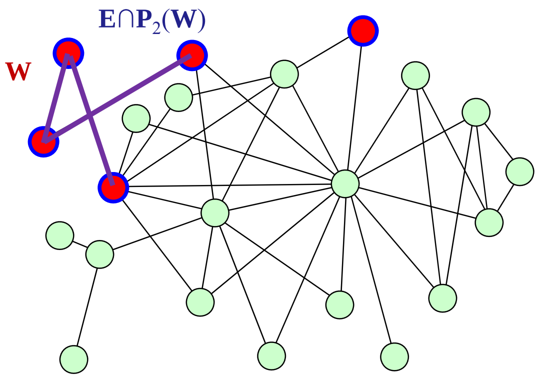


部分グラフ



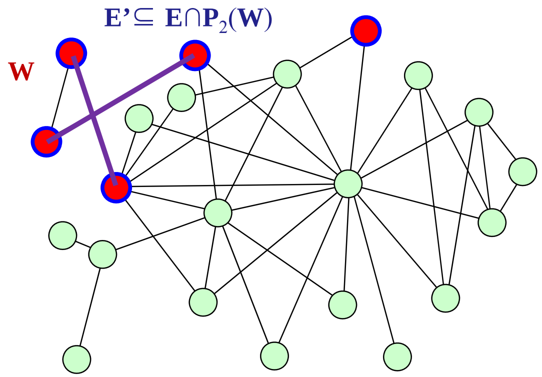
部分グラフ

誘導部分グラフ **G**(**W**)=(**W**, **E**∩**P**$_2$(**W**))



部分グラフ

部分グラフ **H**=(**W**,**E'**)



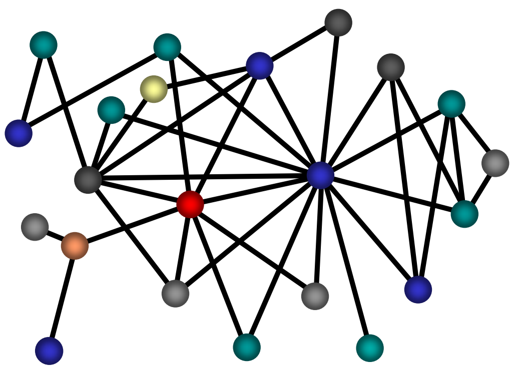
グラフ(11)

連結グラフ（connected graph）

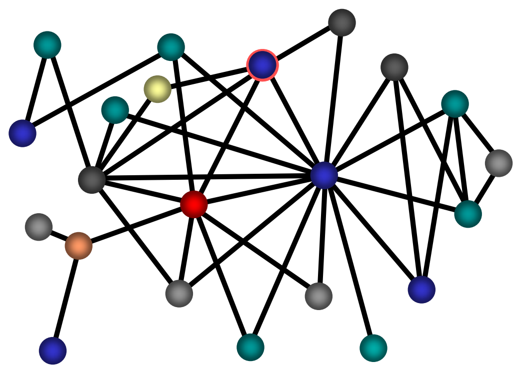
全ての頂点の対v,v'に対して，vからv'へのパスが存在するとき，そのグラフを連結グラフという

連結性

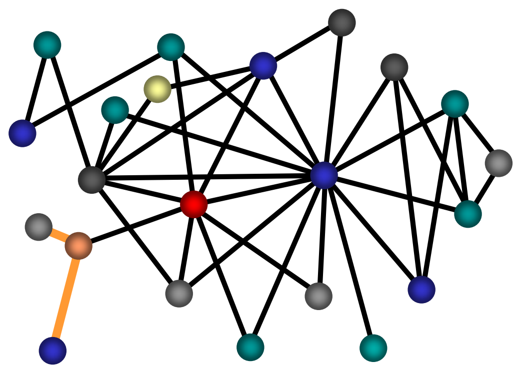
連結度（Connectivity），頑健性（Robustness）



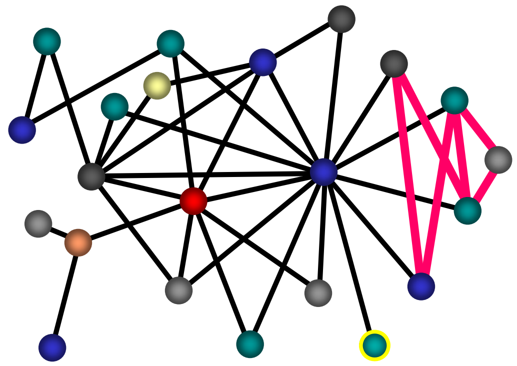
連結度（Connectivity），頑健性（Robustness）



連結度（Connectivity），頑健性（Robustness）



連結度(Connectivity)，頑健性(Robustness)



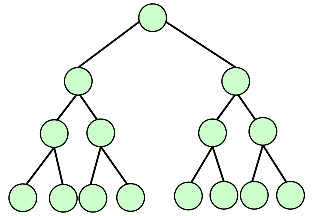
連結性

連結度(connectivity)
Gから何個かの頂点を除去したときに非連結となるか，あるいは孤立点となる．
その最小の個数を連結度と呼び，
$\kappa(G)$
で表す．
先の例では，$\kappa(G) = 1$
K_nでは，$\kappa(G) = n-1$
$\kappa(G) \geqq k$のとき，k-連結という

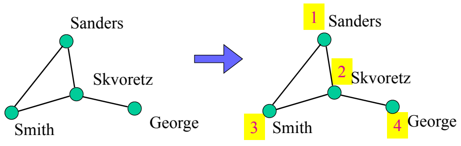
同様に，辺に関して，辺連結度という指標がある．

グラフ(12) 木(tree)

G=(**V**,**E**)が連結であり，サイクルを部分グラフとしてもたないものを木という．
- 下図のように，ルート(root) 根から下方向に表示することが多い
- リンクがある上のノードを親ノード，下を子ノードと呼ぶ
- 木の頂点で次数が1であるものを葉(leaf)という
- 2分木，4分木などがある．



グラフ(13) グラフの表現(データの記述)



dl n = 4, format = nodelist1
labels:
Sanders,Skvoretz,Smith,George
data:
1 2 3
2 1 3 4
3 1 2
4 2

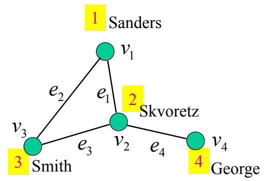
ノードベース

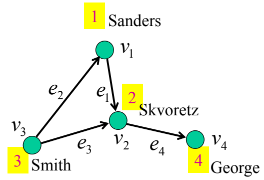
dl n = 4 format = edgelist1
labels:
Sanders,Skvoretz,Smith,George
data:
1 2
1 3
2 1
2 3
2 4
3 1
3 2
4 2

エッジベース

グラフの表現(行列表現:隣接行列1)

Adjacent Matrix

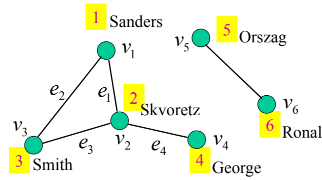




$$\begin{array}{c|cccc} & v_1 & v_2 & v_3 & v_4 \\ \hline v_1 & 0 & 1 & 1 & 0 \\ v_2 & 1 & 0 & 1 & 1 \\ v_3 & 1 & 1 & 0 & 0 \\ v_4 & 0 & 1 & 0 & 0 \end{array}$$

$$\begin{array}{c|cccc} & v_1 & v_2 & v_3 & v_4 \\ \hline v_1 & 0 & 1 & 0 & 0 \\ v_2 & 0 & 0 & 0 & 1 \\ v_3 & 1 & 1 & 0 & 0 \\ v_4 & 0 & 0 & 0 & 0 \end{array}$$

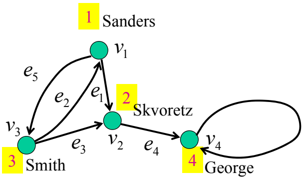
グラフの表現(行列表現:隣接行列2)



行列をA，成分をa_{ij}

$$\begin{array}{c|cccccc} & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \\ \hline v_1 & 0 & 1 & 1 & 0 & 0 & 0 \\ v_2 & 1 & 0 & 1 & 1 & 0 & 0 \\ v_3 & 1 & 1 & 0 & 0 & 0 & 0 \\ v_4 & 0 & 1 & 0 & 0 & 0 & 0 \\ v_5 & 0 & 0 & 0 & 0 & 0 & 1 \\ v_6 & 0 & 0 & 0 & 0 & 1 & 0 \end{array}$$

グラフの表現(行列表現:隣接行列3)

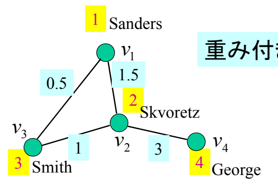


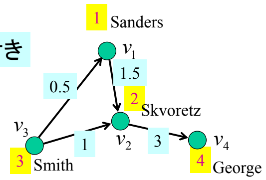
$$\begin{array}{c|cccc} & v_1 & v_2 & v_3 & v_4 \\ \hline v_1 & 0 & 1 & 1 & 0 \\ v_2 & 0 & 0 & 0 & 1 \\ v_3 & 1 & 1 & 0 & 0 \\ v_4 & 0 & 0 & 0 & 1 \end{array}$$

ループと多重辺

グラフの表現(行列表現:隣接行列4)

重み付き





$$\begin{array}{c|cccc} & v_1 & v_2 & v_3 & v_4 \\ \hline v_1 & 0 & 1.5 & 0.5 & 0 \\ v_2 & 1.5 & 0 & 1 & 3 \\ v_3 & 0.5 & 1 & 0 & 0 \\ v_4 & 0 & 3 & 0 & 0 \end{array}$$

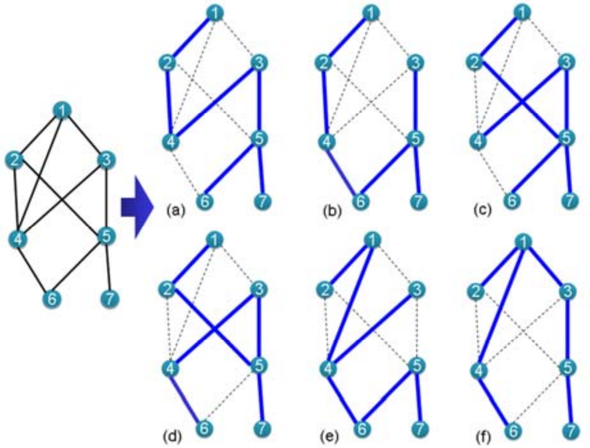
$$\begin{array}{c|cccc} & v_1 & v_2 & v_3 & v_4 \\ \hline v_1 & 0 & 1.5 & 0 & 0 \\ v_2 & 0 & 0 & 0 & 3 \\ v_3 & 0.5 & 1 & 0 & 0 \\ v_4 & 0 & 0 & 0 & 0 \end{array}$$

グラフ(14) 全域木(spanning tree)

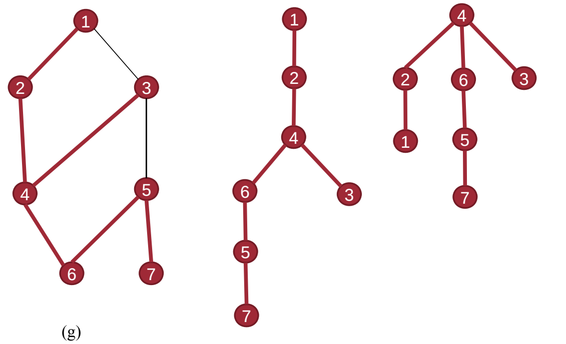
連結なグラフ**G**=(**V**,**E**)に対する全域部分グラフの中で木となるものを全域木という．
- 一つとは限らない．一般には非常に多い

リンクに重みを付けた場合，全域木の中で重みの総計が最小のものを
最小全域木(minimum spanning tree)
という．
- 一つとは限らない．

同様に，最大全域木がある



ルートの選び方は任意



(g)

グラフ(理論)とネットワーク

- 特殊性(一般性)と規模性に違いがある.
  - ノード，リンクの属性を重視するのがネットワーク
  - ネットワークの方が大規模なものを扱う
  - 特に大規模なものは複雑ネットワークと呼ばれることがある.
- 理論的な完結性にも違いがある.
  - 理論的な展開はグラフの方が発展している.
  - グラフの方が，数学的な扱いを重視
    → 接続関係にある種の制限，ルールをおく.
    この意味での特殊性は強い

ここで考える複雑ネットワークにおける,

“大規模性”

とは:

1万ノード程度(以上)

リンク数:数十万程度(以上)

Complex network science

Phenomena and Events on a Web

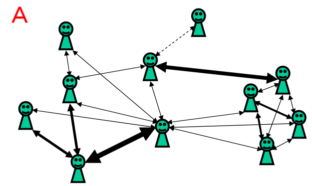
- ◆ 移流拡散に関連する現象・事象
- ◆ 相互作用と創発が鍵となる事象・現象

- 伝染病 → 安心できる生活基盤
- コンピュータウィルス → 健全な情報基盤
- 交通網，物流 → 環境にやさしく，利便性の高い生活基盤
- 電力網 → 環境にやさしく，安心できる生活基盤
- 流行等の経済現象，人間活動 → 持続性のある経済基盤

- 食物連鎖，棲み分けなどの生態系関連 → 環境問題
- 脳内の情報伝達 → 人間を理解，人間の理解(の理解)
- 口コミによる世論形成，広告戦略 → ビジネスモデル
  などなど...

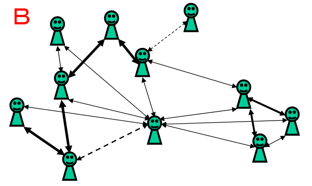
社会学者の視点(興味)

A



- AとBの違いは？
- Aだけで，あるいはBだけで言えることは？

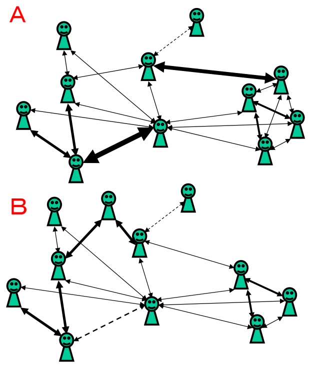
B

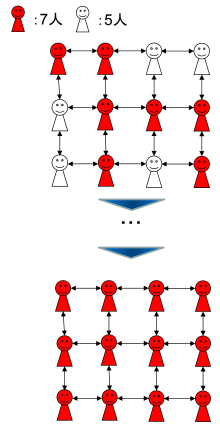


関係性そのものの理解
- ・リンク:紐帯の意味
- ・ノードの位置
- ・構造の変化

など

- AとBの定性的な，さらには定量的な違いは？



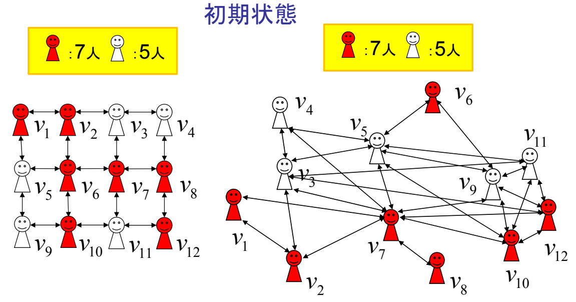


全体の多数決
　赤の意見で全体が決まる.

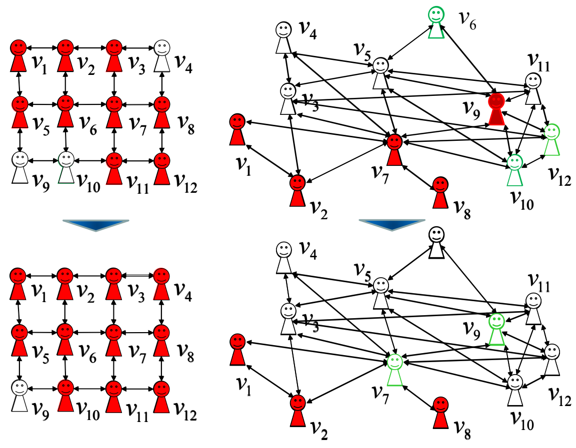
そこからルールを変えても
　赤の意見で全体が決まる.

ある種の合意形成

初期状態



知人(直接つながっている人)の意見で自分の意見を変える
ただし，同数の場合は保留



・ネットワークが存在する(とされる)システムでは，構成要素(例えば人)が**単純に振る舞っても**(単純なルールで行動しても)，ネットワークがシステム全体の振る舞いを**複雑なものにする**.

また，構成要素が**複雑に振る舞っても**システム全体，あるいは一部が**単純に振る舞う場合もある.**

このため,

・例えば，人それぞれを特徴付け，人それぞれの性質を明らかにしても，システム全体の振る舞いを理解することは難しい.

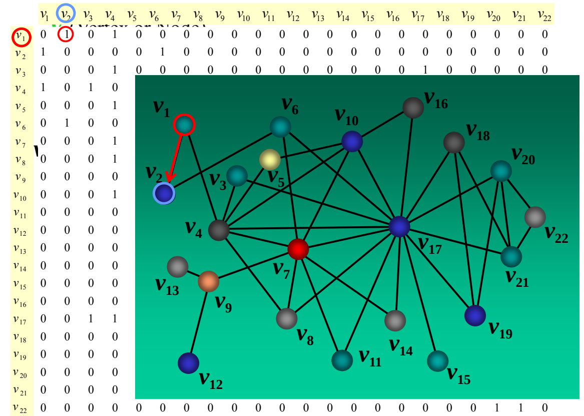
一方,

関係性から生じる**法則性**が垣間見える

そこで,

・ネットワークの性質をより深く理解ことが試みられている.

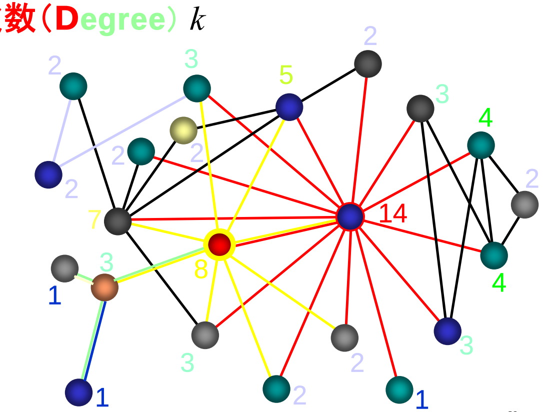
隣接行列



・位数：ノード数 n, N
・サイズ：エッジ(リンク)数 m, M
・ノードの識別子：v_i or i　※ j,kを用いることも
・次数 k_i
　次数分布 $p(k)$
　平均次数 $<k>$
　次数分散 σ^2_n
　最小次数 k_{min}，最大次数 k_{max}
　平均結合相関 knn（隣接ノードの平均次数）
　次数相関 r

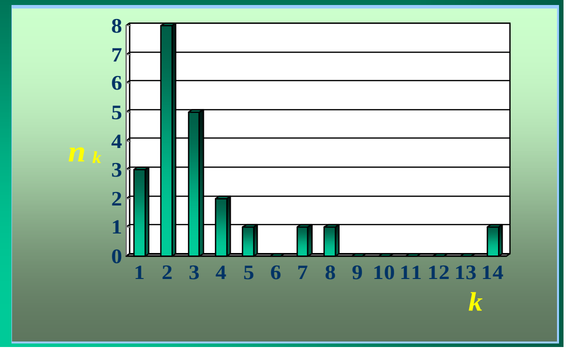
$$r = \frac{4M\sum_{(i,j)\in E} k_i k_j - \left[\sum_{(i,j)\in E}(k_i + k_j)\right]^2}{2M\sum_{(i,j)\in E}(k_i^2 + k_j^2) - \left[\sum_{(i,j)\in E}(k_i + k_j)\right]^2}$$

次数（Degree）k



次数分布 (Degree distribution)

n_k : the number of vertices have degree k

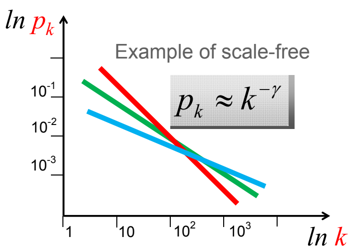


次数分布（Degree distribution）

n : total number of vertices

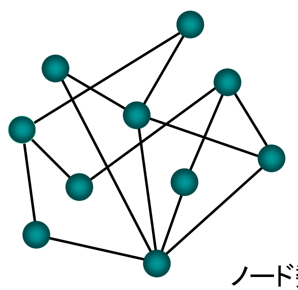
p_k : the probability that a vertex chosen uniformly at random has degree k

$$p_k = \frac{n_k}{n}$$



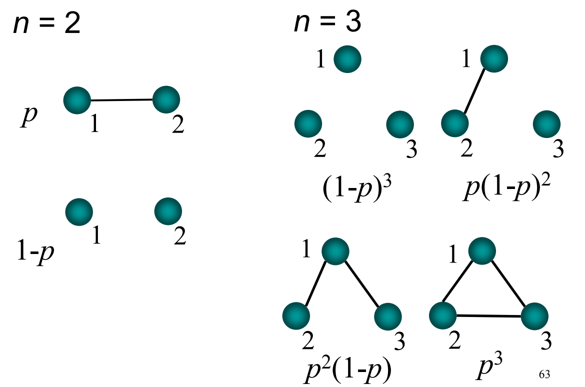
ランダムグラフ(Random Graph)

$k = 3$

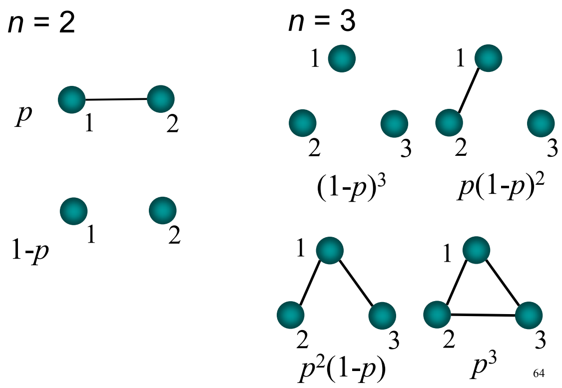


ノード数 n：一定

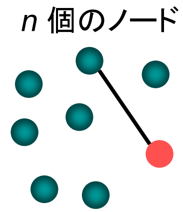
ランダムグラフ(Random Graph)



ランダムグラフ(Random Graph)



ランダムグラフ(Random Graph)



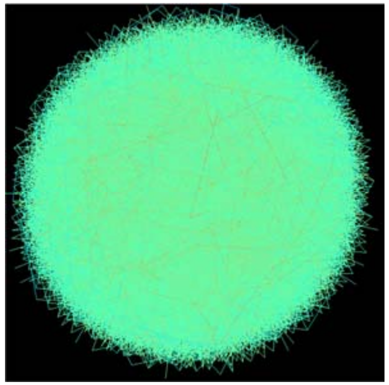
ある 頂点が次数1を持つ確率は？

n-1のノードのうち，1つのノードとリンクを持つ

$$_{n-1}C_1 p^1 (1-p)^{(n-1)-1}$$

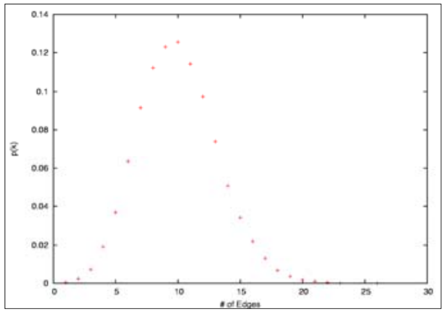
ある 頂点が次数kを持つ確率は？

Random Graph



ランダムグラフ(Random Graph)

$$p(k) = {}_{n-1}C_k p^k (1-p)^{(n-1)-k}$$



ランダムグラフ(Random Graph)

平均次数は？

$$< k >= \sum_{k=0}^{n-1} kp(k)$$

$$= \sum_{k=0}^{n-1} k \cdot {}_{n-1}C_k p^k (1-p)^{(n-1)-k}$$

ランダムグラフ(Random Graph)

二項定理

$$(x+y)^n = \sum_{k=0}^{n} {}_nC_k x^k y^{n-k}$$

$$< k >= (n-1)p$$

ランダムグラフ(Random Graph)

$n \to \infty$のとき

$$\lambda = (n-1)p \quad ※$$

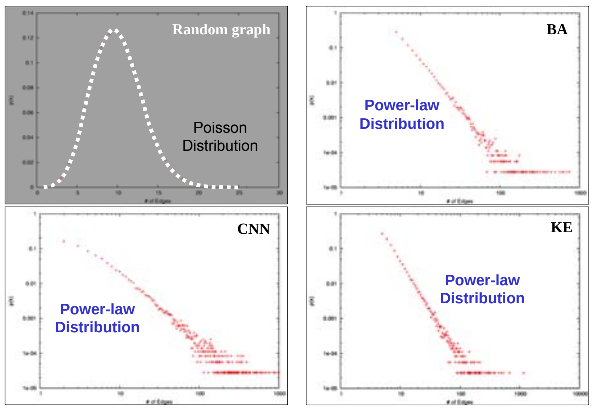
$$\lim_{n\to\infty} {}_{n-1}C_k p^k (1-p)^{(n-1)-k} = e^{-\lambda} \frac{\lambda^k}{k!}$$

$$f(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

ポワソン分布
Poisson Distribution

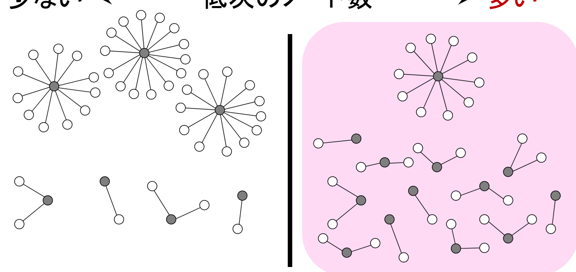
※一般の二項分布の場合は，$\lambda = np$が一定

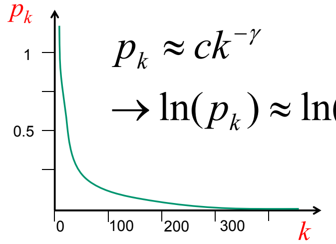
次数分布の例



べき乗指数γ

小さい	← べき乗指数γ →	大きい
多い	← 高次のノード数 →	少ない
少ない	← 低次のノード数 →	多い





$$p_k \approx ck^{-\gamma}$$

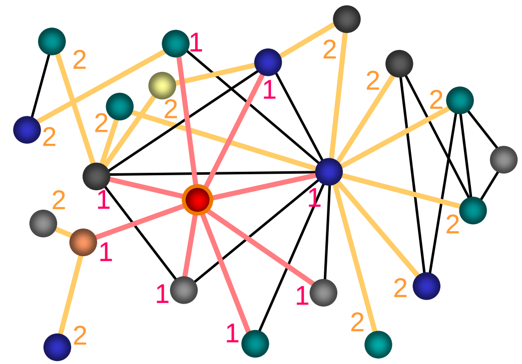
$$\to \ln(p_k) \approx \ln(c) - \gamma \ln(k)$$

スケールフリー

$$p_k \approx c(sk')^{-\gamma}$$ sは代表スケール

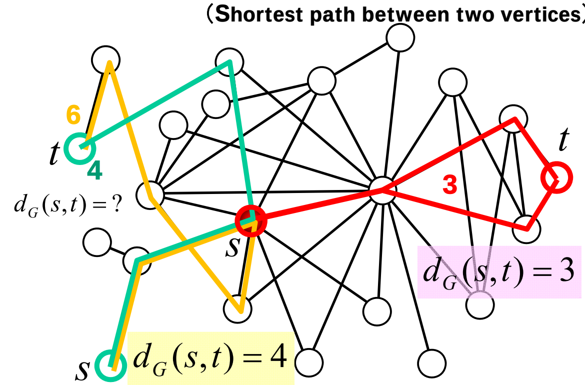
$$\to \underline{\ln(p_k)} \approx \ln(c) - \gamma\ln(s) \underline{- \gamma \ln(k')}$$

グラフ距離 Distance on a Network



頂点間距離，最短距離 $d_G(s,t)$

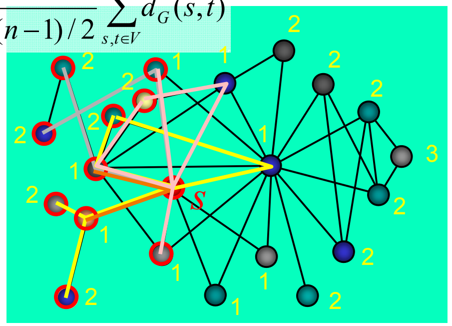
(Shortest path between two vertices)



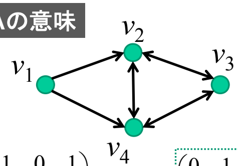
平均頂点間距離

Mean geodesic (i.e., shortest) distance

$$L = \frac{1}{n(n-1)/2} \sum_{s,t \in V} d_G(s,t)$$



AxAxAx…Aの意味



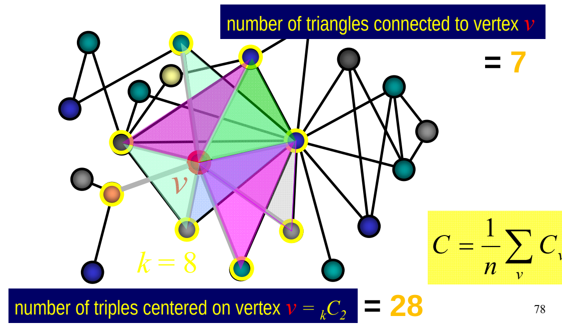
$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \qquad \mathbf{A}^2 = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

$$\#(v_2 \to v_* \to v_2) \quad a_{22}^{(2)} = a_{21}a_{12} + a_{22}a_{22} + \underline{a_{23}a_{32}} + \underline{a_{24}a_{42}} = 0\cdot1 + 0\cdot0 + 1\cdot1 + 1\cdot1 = 2$$

$$\#(v_2 \to v_* \to v_3) \quad a_{23}^{(2)} = a_{21}a_{13} + a_{22}a_{23} + a_{23}a_{33} + \underline{a_{24}a_{43}} = 0\cdot0 + 0\cdot1 + 1\cdot0 + 1\cdot1 = 1$$

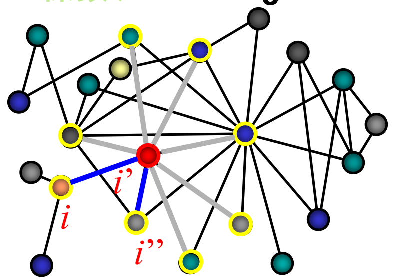
クラスター係数（Clustering coefficient）

$$C_v = \frac{\text{number of triangles connected to vertex } v}{\text{number of triples centered on vertex } v}$$



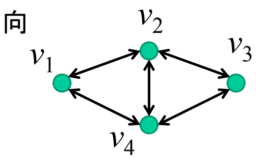
$$C = \frac{1}{n}\sum_v C_v$$

クラスター係数（Clustering coefficient）



$$C = \frac{\text{分母の内，}i\text{と}i''\text{が隣接してるものの数}}{(i,i')\text{と}(i',i'')\text{がエッジである}i,i',i''\text{の数}}$$

無向



$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

$$a_{11}^{(2)} = a_{11}a_{11} + a_{12}a_{21} + a_{13}a_{31} + a_{14}a_{41} = k_1$$

$$\mathbf{A}^3 = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}\begin{pmatrix} 2 & 1 & 2 & 1 \\ 1 & 3 & 1 & 2 \\ 2 & 1 & 2 & 1 \\ 1 & 2 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 5 & 2 & 5 \\ 5 & 4 & 5 & 5 \\ 2 & 5 & 2 & 5 \\ 5 & 5 & 5 & 4 \end{pmatrix}$$

三角形の数 = $\frac{1}{6} tr(\mathbf{A}^3)$ = 2

Small world & Scale free

Small World　L:小さい
C:大きい

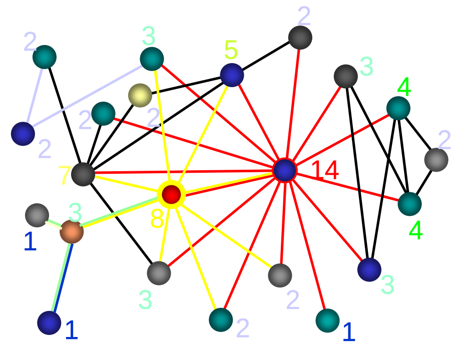
Scale Free
次数の不均一性：ベキ乗則

Centrality（中心性）

どの要素が重要？

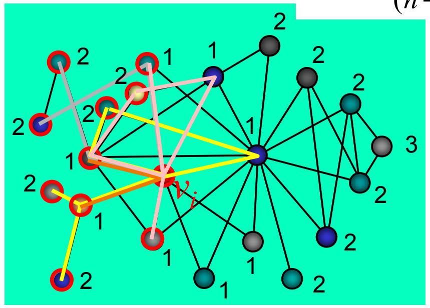
中心性指標 （Centrality indices）

次数中心性 （Degree Centrality）



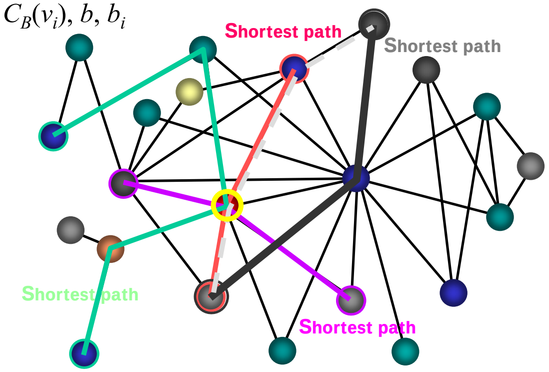
近接中心性 （Closeness Centrality）

$$C_C(v_i) = \frac{1}{L_i} \qquad L_i = \frac{\sum_{t \in V, t \neq v_i} d_G(v_i, t)}{(n-1)}$$



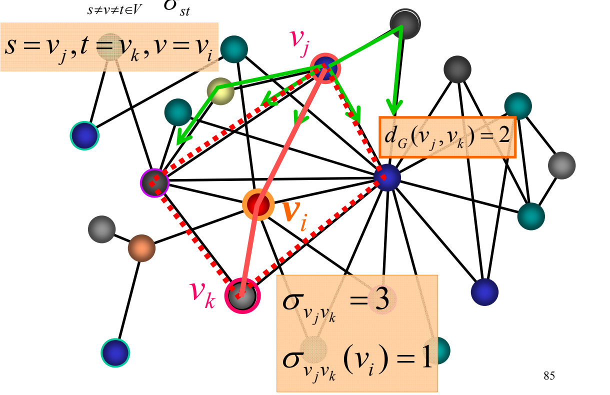
媒介中心性 （Betweeneess Centrality）

$C_B(v_i)$, b, b_i



$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Betweeneess Centrality



固有ベクトル中心性

$$\mathbf{u}^{(k+1)} = \mathbf{A}\mathbf{u}^{(k)}$$

$$u_i = \frac{1}{\lambda} \sum_{j}^{N} a_{ij} u_j$$

ボナチッチ（1972）

Perron-Frobeniusの定理

Aが非負行列（$a_{ij} \geqq 0$）

(1) Aは非負の固有値をもつ

λ$_1$を最大固有値，$\boldsymbol{u}_1$を対応する固有ベクトル

$\boldsymbol{u}_1$は非負のベクトル

(2) Aは既約（分解不能）なら，λ$_1$は重複度1

グラフが連結ならば**A**は既約（分解不能）

固有ベクトル中心性

$$\mathbf{A}\mathbf{e}_j = \lambda_j \mathbf{e}_j$$

簡単のため，*N*個の固有ベクトルは一次独立とする．

固有値が重複しているときは広義固有ベクトルを使う．

$$(\mathbf{A} - \lambda_k \mathbf{I})^p \mathbf{e}_k = 0$$

$$\mathbf{u}^{(0)} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + \cdots + c_N\mathbf{e}_N$$

$$\mathbf{A}\mathbf{u}^{(0)} = c_1\mathbf{A}\mathbf{e}_1 + c_2\mathbf{A}\mathbf{e}_2 + \cdots + c_N\mathbf{A}\mathbf{e}_N$$

$$= c_1\lambda_1\mathbf{e}_1 + c_2\lambda_2\mathbf{e}_2 + \cdots + c_N\lambda_N\mathbf{e}_N$$

固有ベクトル中心性

$$\mathbf{u}^{(0)} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + \cdots + c_N\mathbf{e}_N$$

$$\mathbf{u}^{(k)} = \mathbf{A}^k\mathbf{u}^{(0)} = c_1\lambda_1^k\mathbf{e}_1 + c_2\lambda_2^k\mathbf{e}_2 + \cdots + c_N\lambda_N^k\mathbf{e}_N$$

$$\mathbf{u}^{(k)} = \lambda_1^k[c_1\mathbf{e}_1 + c_2(\frac{\lambda_2}{\lambda_1})^k\mathbf{e}_2 + \cdots + c_N(\frac{\lambda_N}{\lambda_1})^k\mathbf{e}_N]$$

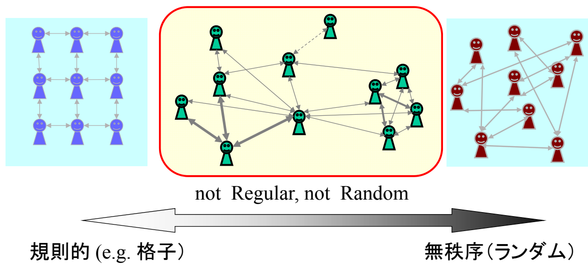
$$\mathbf{u}^{(k)} \cong c_1\lambda_1^k\mathbf{e}_1$$

現実のネットワークを表すために考えられた数理モデル

Network Models

ネットワークを作り，知る

現実のネットワーク構造は？



- 人的（社会的）ネットワークの場合，完全に規則的でも完全に無秩序でもない複雑な構造をしているらしい
- 人的以外のネットワークでも同様であることが示されている．

システムのモデリング（ネットワーク）

製品を構成する部品間の関係性などは後述する.

製品を構成する部品間の関係性などの直接的なもの以外でも

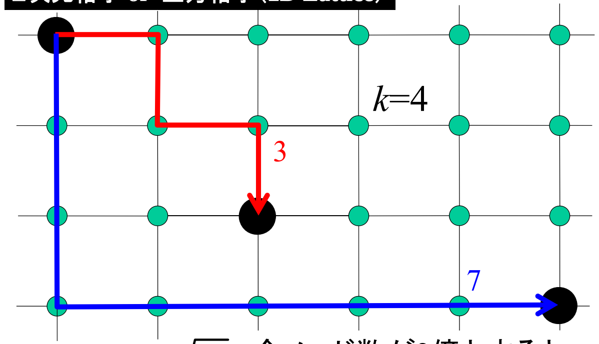
- ・口コミ，流行の爆発
- ・伝染病の蔓延
- ・イノベーションの起源，そして発展
- ・企業の発展モデル
- ・企業，地域クラスターの形成
- ・交通網，物流

このようなものをシステムとして考える場合に，背後のネットワークをモデル化すると見通しがよくなることが多い.

- **WS (Watts-Strogatz) Model** [Watts, Strogatz1998]
  - ネットワークの狭さと高い凝集性を再現.
- **BA (Barabási-Albert) Model** [Barabási, Albert1999]
  - ネットワークの成長過程をモデル化. "Rich gets Richer"モデル.
  - 次数分布の冪乗則とネットワークの狭さを再現.
- **KE (Klemm-Eguíluz) Model** [Klemm, Equíluz2002]
  - ノードに活性・非活性の概念を導入した成長モデル.
  - 次数分布の冪乗則と高い凝集性を再現.
- **CNN (Connecting Nearest Neighbor) Model** [Vásquez2003]
  - 局所規則に基づいてノードとエッジの生成規則を定める成長モデル.
  - 次数分布の冪乗則と高い凝集性を再現.
- **Fitness / Threshold Model** [Cardareli2000など]
  - ノードに重みを与え，それによってエッジの生成規則を定める成長モデル.
  - 次数分布の冪乗則とネットワークの狭さを再現.

・・・など多くが提案されている. ネットワークの構造やその生成規則をモデル化している.

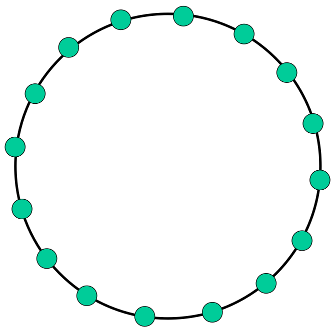
2次元格子 or 正方格子（2D Lattice）



$$L \propto \sqrt{n}$$

全ノード数が2億とすると，Lは10000のオーダ

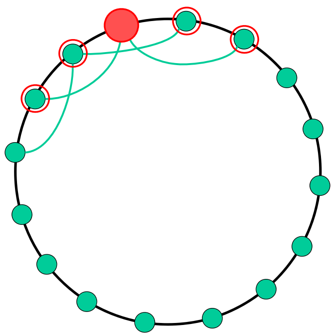
1次元格子（1D Lattice）



$$L \propto n$$

$$C = 0$$

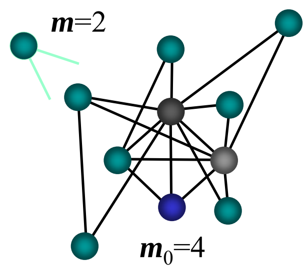
1次元格子（1D Lattice）に規則的にリンクを追加



$$L \propto \frac{n}{2}$$

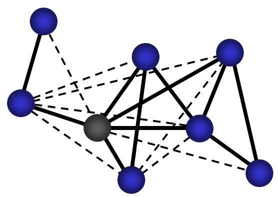
$$C = \frac{3}{{}_4C_2} = 1/2$$

BA (Barabási-Albert)



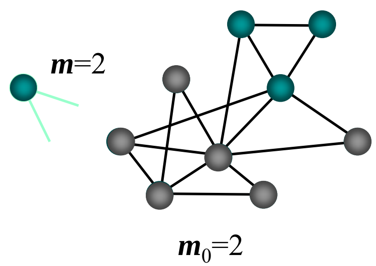
友人が沢山いる人は，更に友人が増える

CNN (Connecting Nearest Neighbor)



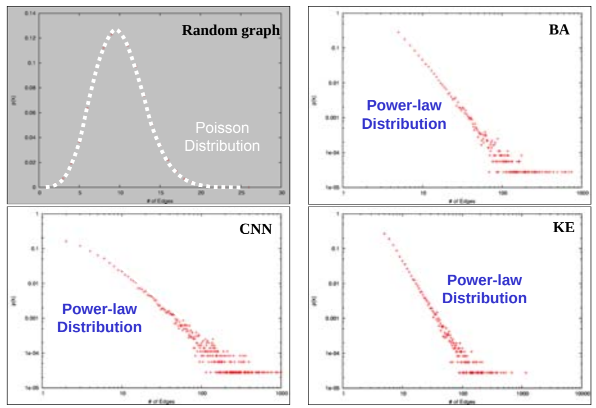
友人の知人は，友達になりやすい

KE (Klemm-Eguíluz Model)



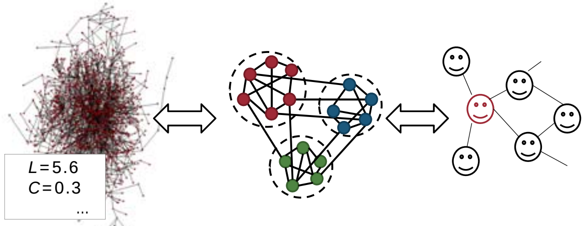
友人の知人は，友達になりやすいのだが，時々，人間関係に疲れる人がいる

次数分布



ネットワークの階層構造 (i)

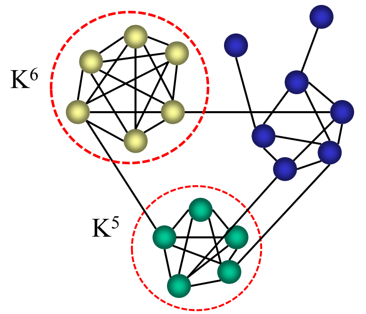
大局的構造　　コミュニティ構造※　　局所的構造※※



※ 複雑ネットワーク科学では，位相構造から抽出される部分グラフを
“コミュニティ”と”クラスター”の両方で呼ぶことが多い．
構造という観点では，メソ構造がふさわしい言い方かもしれない．
※※　局所情報のみから得られる構造と，大局的な情報を用いて得られるが個々を識別できる要素を集めた
構造の2つを指す．

階層性（構造）　▶　完全グラフの内包

クリークの抽出（社会科学）

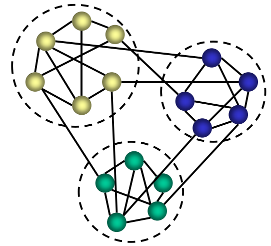


階層性（構造）　▶　コミュニティ構造

局所的に密集したサブグラフ構造

[Newman2004]

⇒ 社会的ネットワークの場合，互いに同質な集団を意味していることが多い．



・Modularity Q　コミュニティ構造→モジュラー構造

$$Q=\sum_{\alpha=1}^{K}\left[\frac{l_\alpha}{m}-\left(\frac{d_\alpha}{2m}\right)^2\right]$$

l_α: α内のリンクの数，d_α: α内の次数の合計

$$e_{\alpha\beta}=\frac{1}{2m}\sum_{i\in\alpha}\sum_{j\in\beta}a_{ij},$$

$$e_{\alpha\alpha}=\frac{1}{2m}\underline{\sum_{i\in\alpha}\sum_{j\in\alpha}}a_{ij}=\frac{l_\alpha}{m}$$

$$e_{\alpha\beta}=\frac{1}{2m}\sum_{i\in\alpha}\sum_{j\in\beta}a_{ij}$$

コミュニティαとβを結ぶリンク数のリンク総数(m)の2倍※に対する割合

※無向グラフの場合，一つのリンクは二重にカウントされるので2倍のリンク数に対する割合とする．

$$a_\alpha=\sum_{\gamma}e_{\alpha\gamma}=\frac{d_\alpha}{2m}$$

・a_αは，ランダムに選ばれたノードがコミュニティα内のノードとリンクされている確率．
・ランダムに選ばれたノードがコミュニティα内のノードである確率もa_α

a_α^2：ランダムに選んだ2つのノードがコミュニティαに存在する確率

同じコミュニティ内のノード同士を結ぶリンクの割合$e_{\alpha\alpha}$がa_α^2に近ければネットワークがランダムグラフに近い．

$$Q_\alpha=e_{\alpha\alpha}-a_\alpha^2$$
$$Q=\sum_\alpha Q_\alpha$$

・ネットワーク全体が一つのコミュニティのとき：

$$Q=0$$

・すべてのノードが異なるコミュニティになる場合：

$e_{\alpha\alpha}=0$，$a_\alpha=k_\alpha/m$より，

$$Q=\sum_\alpha Q_\alpha=-\sum_\alpha\left(\frac{k_\alpha}{m}\right)^2<0$$

コミュニティ1内のノード同士を結ぶリンクの数をm_1，
コミュニティ2内のノード同士を結ぶリンクの数をm_2，
コミュニティ1内とコミュニティ2内を結ぶリンクの数をm_{12}

$$e_{11}=\frac{m_1}{m},e_{12}=\frac{m_{12}}{2m},e_{21}=\frac{m_{12}}{2m},e_{22}=\frac{m_2}{m},$$

$$a_1=\frac{2m_1+m_{12}}{2m},$$

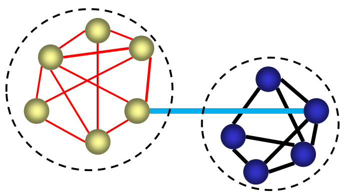
$$a_2=\frac{m_{12}+2m_2}{2m},$$

$$Q_1=e_{11}-a_1^2,$$

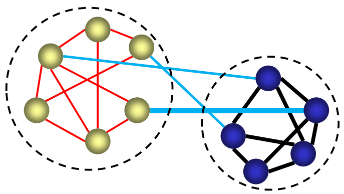
$$Q_2=e_{22}-a_2^2,$$

$$Q=Q_1+Q_2=\frac{1}{2m^2}\{4m_1m_2-m_{12}^2\}$$

$m_1, m_2>>m_{12}$
$Q>0$



$m=20$
$m_1:11,$
$m_2:8,\ m_{12}:1$
$Q=351/800$



$m=20$
$m_1:9,$
$m_2:8,\ m_{12}:3$
$Q=285/800$