アルゴリズムとプログラミング実践講座

http://akashi.ci.i.u-tokyo.ac.jp/mary/lectures/algorithm/

火曜　13:00 -- 14:30
I-REF 棟 6階　ヒロビー

稲葉真理　with 浅井大史・手塚宏史

先週のレポートから(WS)

平均次数は4 ,
ノード数をN, 繋ぎかえるエッジの割合をP
各ノードを、円周上に配置した場合の例

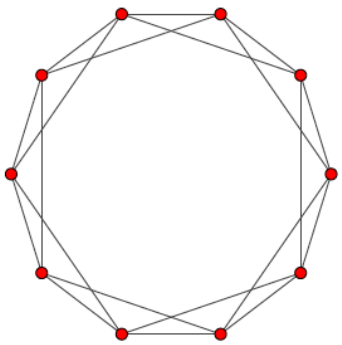


図 1　WS モデル $(N = 10, P = 0)$

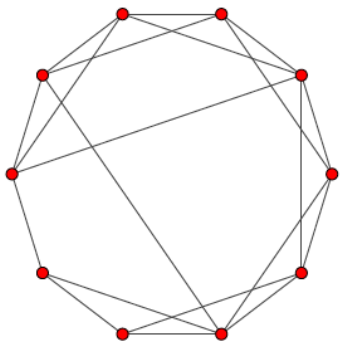


図 2　WS モデル $(N = 10, P = 0.1)$

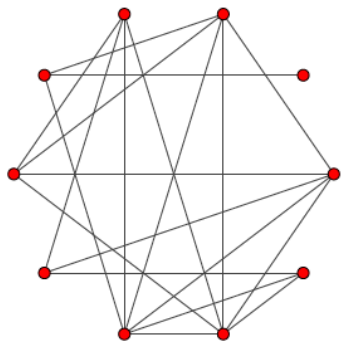


図 3　WS モデル $(N = 10, P = 1.0)$

先週のレポートから(WS)

平均次数は4 ,
ノード数をN, 繋ぎかえるエッジの割合をP

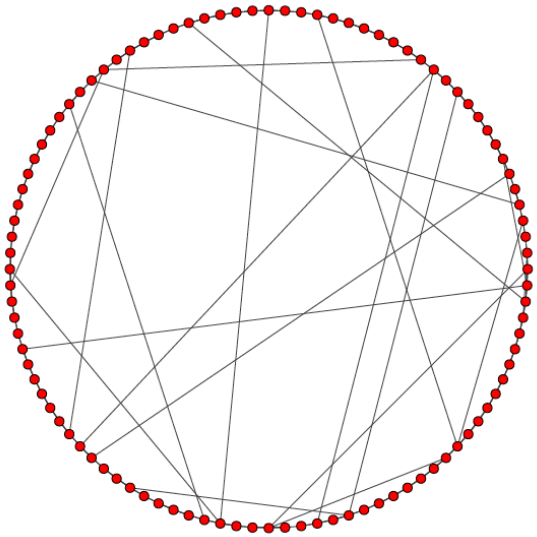


図 4　WS モデル ($N = 100, P = 0.1$)

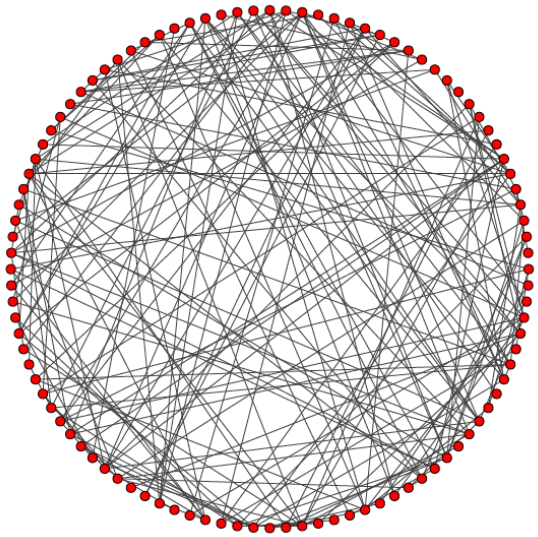


図 5　WS モデル ($N = 100, P = 1.0$)

先週のふりかえり

情報があふれてる世界で

（1） 探す

– 欲しい情報を探す （Information Retrieval)

- data mining, クラスタリング，機械学習, 統計

（Google という名の広告屋さんは、とても儲かっている♪）

（2） 予測する (complex network)

– 情報がどう流れるか？

- **どう広告をうつのが有効か？ インフルエンサー**

先週のふりかえり

現実の世界を反映するような
数理モデルが欲しい

WSモデル [1998]
（「スモールワールド」序文より）

Watts が父から聞いた話

「おまえはたかだか６人の人を介すればアメリカ合衆国大統領と握手できるんだよ」

「いかなる二人であっても数人を介して握手できる世界とはどのようなものなのか、どのようなダイナミクスが隠されているのかを明らかにするという研究テーマは？」博士論文執筆中の **Watts** は　**Strogatz** 先生に意見を求めた

現実世界の（役に立つ）
ネットワークの3つの特徴
（これも多数決　←　便利と思う人が多い）

- スモールワールド性
  → 任意の2ノードの平均距離は短い

- スケールフリー性
  → ノードの次数がベキ乗則

- クラスター性
  → あるノードと、隣接ノードの隣接ノードとの間は
  エッジがある確率が高い

８０：２０の法則 （パレートの法則）
[ヴィルフレド・パレート (1848-1923)]

経済学者・社会学者
「法則」というより、理系的には、むしろ「観察」

世の中の富の 80% は、20% の金持ちが所有し、
残りの20%の富を、80%の貧乏人で分けている

→ べき乗則の発見

べき乗則

たとえば　$y = (1/x)^{\alpha}$ のような関数　(α=1, 2, …)

たとえば、上に k 倍に引き延ばしても、
横にずらすとグラフが重なる

→ スケールフリー

上の式の両辺　ログとると　$\log y = -\alpha \log x$

→　x、y　両方を対数軸にするとグラフが直線

でも・・・

- ちょっと古い教科書を見ると、

「指数分布」は良くみるけれど、

「べき分布」は、それほどでてこない。

→　これも多数決

[A.L. バラバシ, 2000]

WWW のリンクの数を数えて考察した

インターネットは、ランダムネットワークではない。
インターネットの「リンク」は、べき乗則に従う。
ノードが増えても、構造がかわらない

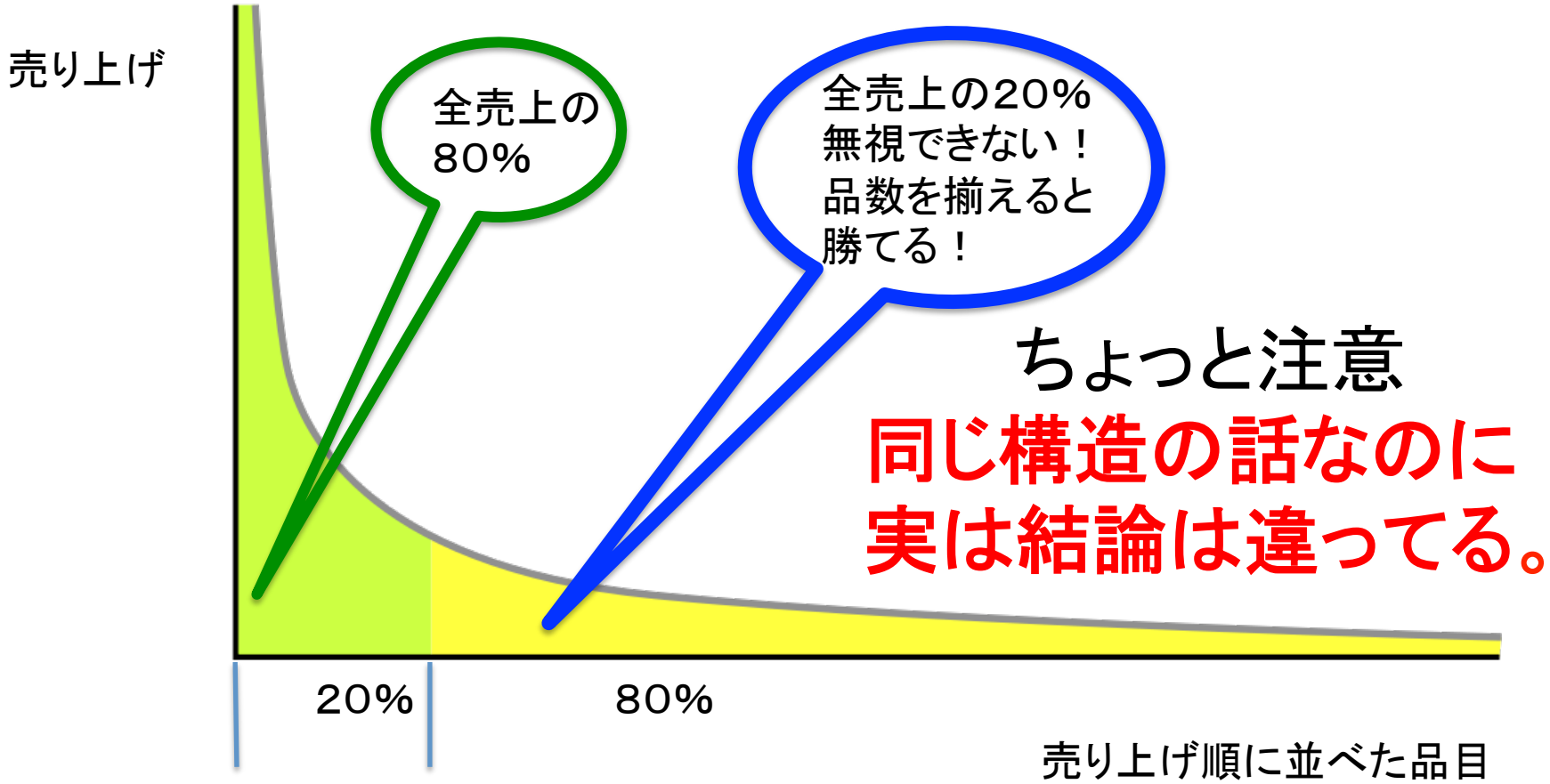
→ 「スケールフリー」と特徴づけた

バラバシ・アルバートモデル （BA モデル）
スケールフリー性を持つモデルを実現

べき乗則

選択と集中

amazon など　Long　Tail



ちょっと注意

同じ構造の話なのに
実は結論は違ってる。

グラフは、http://ja.wikipedia.org/wiki/ロングテール　より

スケールフリー（と言われてる）ネットワーク

- WWW（ウェブページとリンク）
- インターネット（ルータとその接続）
- 学術論文（エルデシュ数）

→ ノードを増やしても、構造がかわらない

→ （ランダムネットワークに比べ）故障に強い

→ （ランダムネットワークに比べ）攻撃に弱い

現実世界の（役に立つ）ネットワークの特徴

ネットワーク	ノード数	平均次数	ベキの指数
www	203549046	10.46	2.1/2.7
インターネットルータ	150000	2.7	2.3
論文引用	783339	8.57	2.5
航空路線網	3880	9.7	3
単語同時出現	478773	74.3	2.7
俳優共演関係	449913	113.443	2.3

「複雑ネットワークとその構造(矢久保考介著)」より

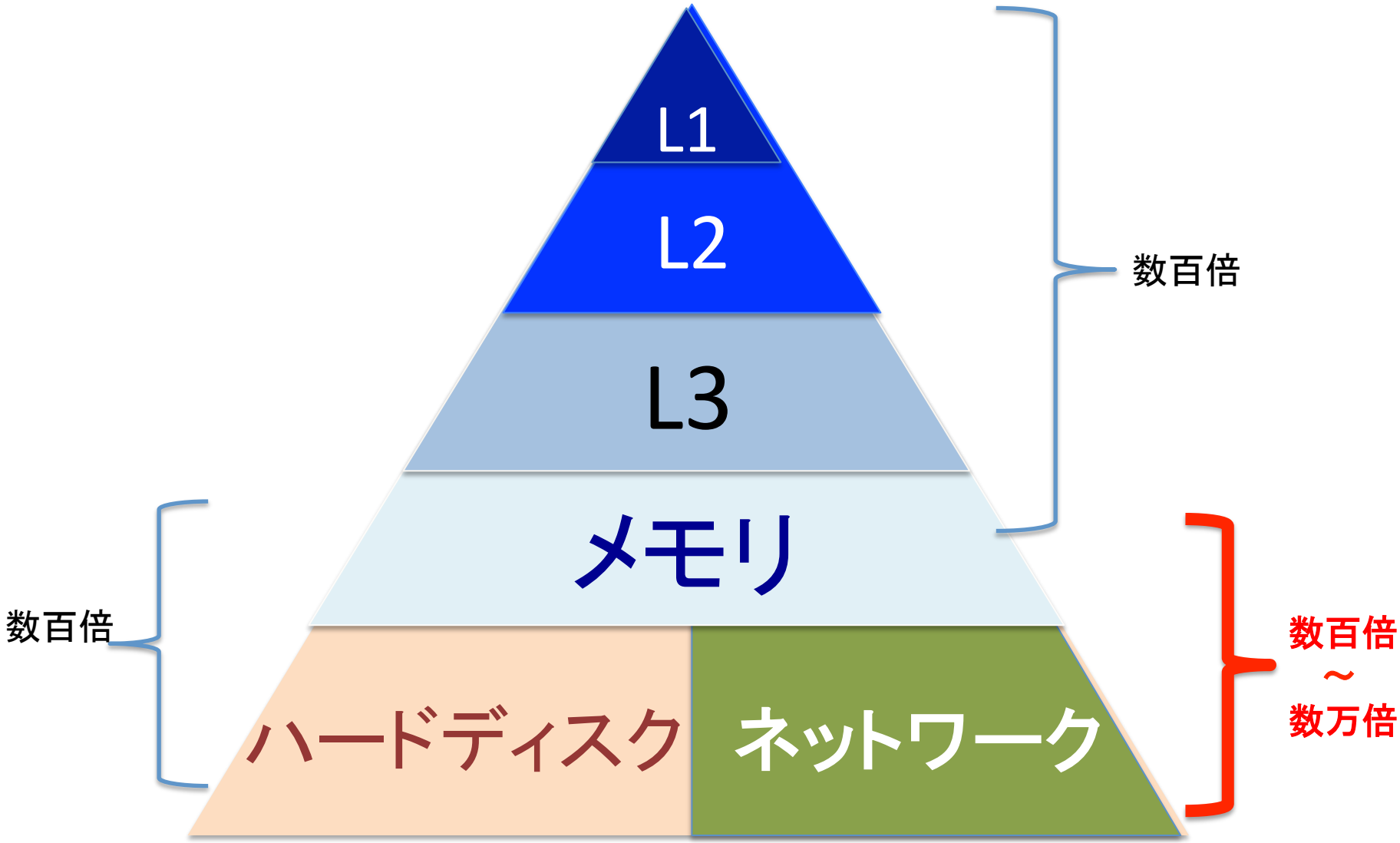
ネットワークのモデル

1990年代後半から、急速に進展中

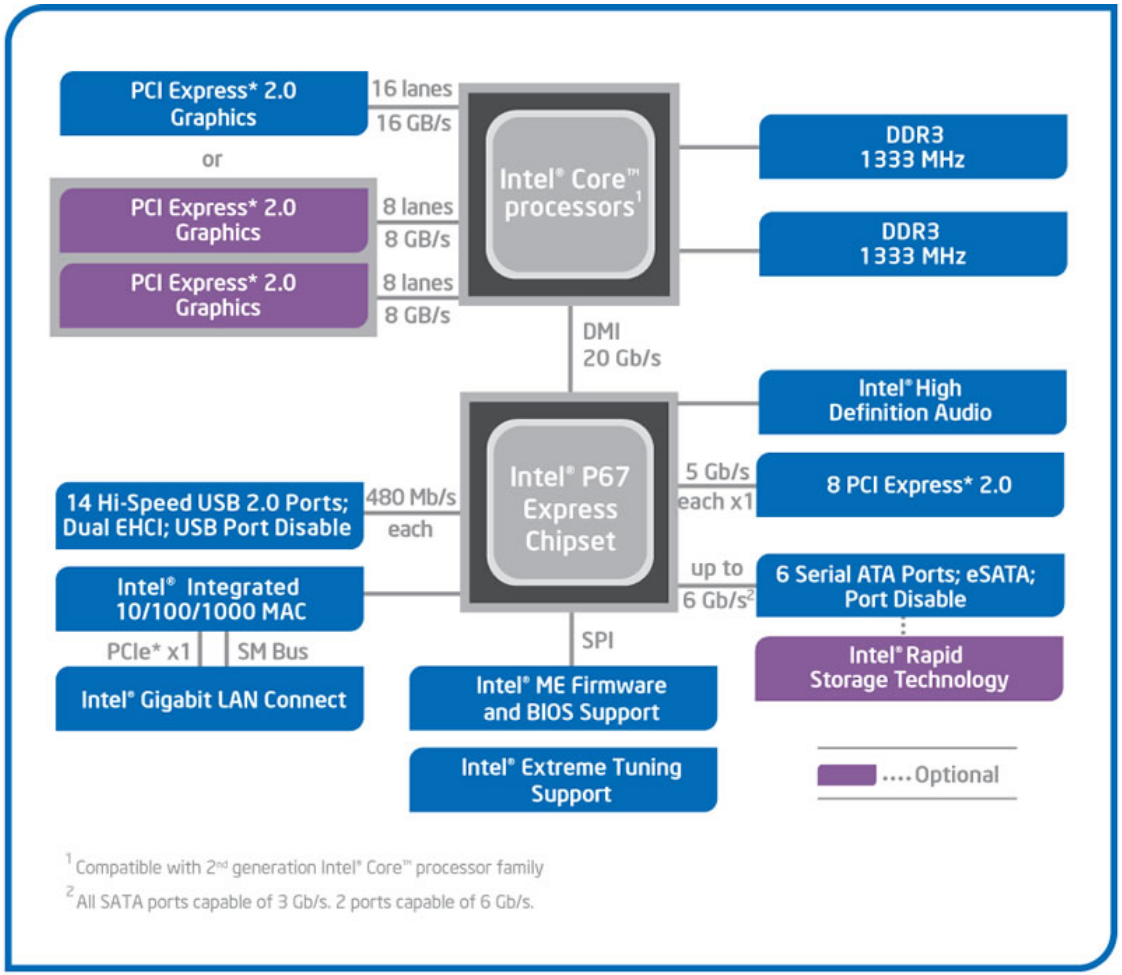
→　広告をどううつべきか（Google など）
→　通販の在庫をどうすべきか？(Amazonなど)

先の3つ以外にも、
色々なバリエーションがでつつある。
　→　新規参入組も、多数決で生き残りが決まる♪

（おまけ）メモリー? 階層



例：チップセット

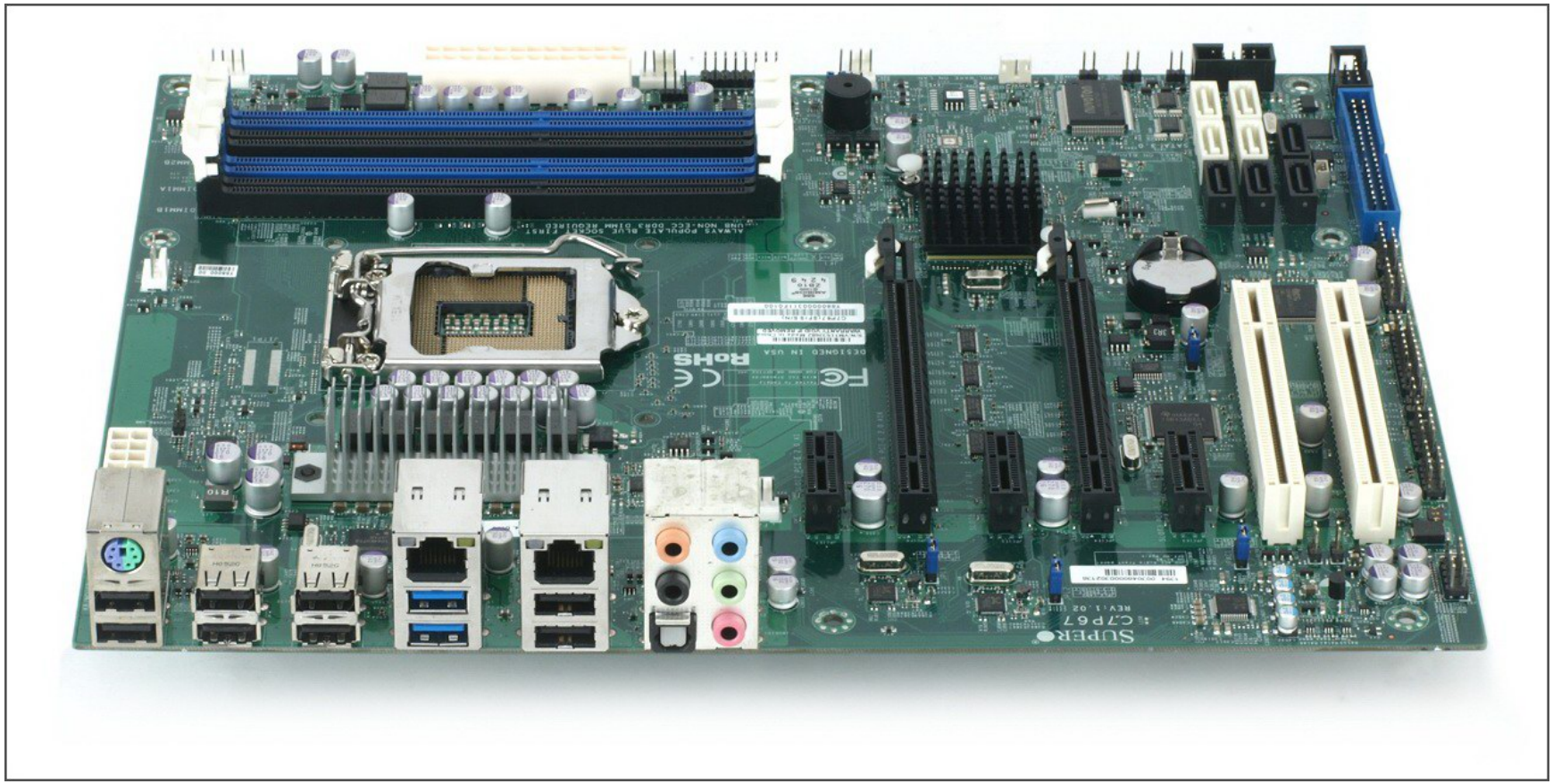


Intel® P67 Express Chipset Platform Block Diagram

http://www.intel.com/content/www/us/en/chipsets/mainstream-chipsets/p67-express-chipset.html

例えば　チップセット　P67

例えばマザーボード
Supermicro C7P67



http://tweakers.net/pricewatch/291321/supermicro-c7p67.html

Ethernet Card

http://www.chelsio.com/wp-content/uploads/2013/10/T520-CR-PB.pdf

（このカードはPCI Express 3 にさす、１OGbps　カード）

色々なイーサネットの誕生年(?)と速さ

1983 IEEE 802.3 10BASE5

1990 IEEE 802.3i 10BASE-T

1994 IEEE 802.3u 100BASE-T

1998 IEEE 802.3z 1000BASE-SX

2003 IEEE 802.3ae 10Gbps Ethernet over fiber

2010 IEEE 802.3ba 40Gbps, 100Gbps

もちろん、もっといろいろあります

規格名		別名	通信速度	標準化規格	使用ケーブル	距離
1BASE5			1Mbps	IEEE 802.3e	UTP（2対）	250m
10BASE2		Thin Ethernet	10Mbps	IEEE 802.3a	50Ω同軸 (5mm)	185m
10BASE5		Thick Ethernet		IEEE 802.3	50Ω同軸 (12mm)	500m
10BASE-T				IEEE 802.3i	UTP/STP (Cat3)	100m
10BROAD36				IEEE 802.3b	75Ω同軸	3600m
10BASE-F	10BASE-FB			IEEE 802.3j	光マルチモード	2000m
	10BASE-FP				光マルチモード	1000m
	10BASE-FL				光マルチモード	2000m
100BASE-T	100BASE-TX	Fast Ethernet	100Mbps	IEEE 802.3u	UTP (Cat5)	100m
	100BASE-T4				UTP（4対Cat3）	100m
	100BASE-T2			IEEE 802.3y	UTP（2対Cat3）	100m
100BASE-F	100BASE-FX			IEEE 802.3u	光マルチモード/シングルモード	2000m/20km
1000BASE-T	1000BASE-T	Gigabit Ethernet	1000Mbps	IEEE 802.3ab	UTP（4対Cat5e）	100m
	1000BASE-TX			TIA-EIA/-854	UTP（4対Cat6）	100m
1000BASE-X	1000BASE-SX			IEEE 802.3z	光マルチモード	550m
	1000BASE-LX				光マルチモード/シングルモード	550m/5000m
	1000BASE-CX				同軸ケーブル（2芯並行）	25m
10GBASE-T			10Gbps	IEEE 802.3an	UTP（4対Cat6）	100m
10GBASE-R	10GBASE-SR			IEEE 802.3ae	光マルチモード	300m
	10GBASE-LR				光シングルモード	10km
	10GBASE-ER				光シングルモード	40km
	10GBASE-ZR				光シングルモード	40km以上を想定
10GBASE-W	10GBASE-SW				光マルチモード	300m
	10GBASE-LW				光シングルモード	10km
	10GBASE-EW				光シングルモード	40km
10GBASE-X	10GBASE-LX4				光シングルモード	10km
10GBASE-CX				IEEE 802.3ak	4対2芯銅線 (CX4) 同軸	15m
40GBASE-R	40GBASE-KR4		40Gbps	IEEE 802.3ba		
	40GBASE-CR4					
	40GBASE-SR4					
	40GBASE-LR4					
100GBASE-R	100GBASE-CR4		100Gbps	IEEE 802.3bj		
	100GBASE-CR10			IEEE 802.3ba		
	100GBASE-SR10					
	100GBASE-LR4					
	100GBASE-ER4					

http://ja.wikipedia.org/wiki/イーサネット#.E6.AD.B4.E5.8F.B2

驚きなのは

1983 IEEE 802.3 10BASE5

1990 IEEE 802.3i 10BASE-T

割と普通に使える

1994 IEEE 802.3u 100BASE-T

1998 IEEE 802.3z 1000BASE-SX

2003 IEEE 802.3ae 10Gbps Ethernet over fiber

2010 IEEE 802.3ba 40Gbps, 100Gbps

（おまけ）ネットワーク

データの転送速度を決めるもの

→　内部バスのスピード

→　ネットワークインターフェースのスピード

→　他にも　TCP　とか UDP とか・・・

現実世界の（役に立つ）
ネットワークの3つの特徴
（これも多数決　←　便利と思う人が多い）

- スモールワールド性
  → 任意の2ノードの平均距離は短い

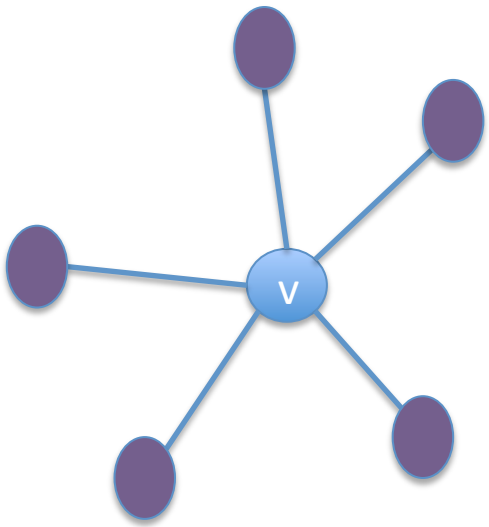
- スケールフリー性
  → ノードの次数がベキ乗則

- クラスター性
  → あるノードと、隣接ノードの隣接ノードとの間は
  エッジがある確率が高い

クラスター係数

「自分の友達と、自分の友達は、けっこう、友達同士のことが多いよね？」

ある頂点に着目



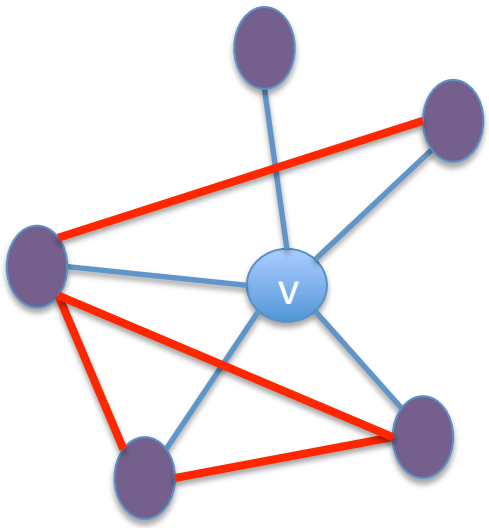
頂点　v のクラスター係数 Cv

$$\frac{\text{vの隣接ノード間のエッジの総数}}{\text{完全グラフの場合のエッジ数}}$$

クラスター係数

「自分の友達と、自分の友達は、けっこう、友達同士のことが多いよね？」

ある頂点に着目



頂点　v のクラスター係数 Cv

$$= \frac{\text{vの隣接ノード間のエッジの総数}}{\text{完全グラフの場合のエッジ数}}$$

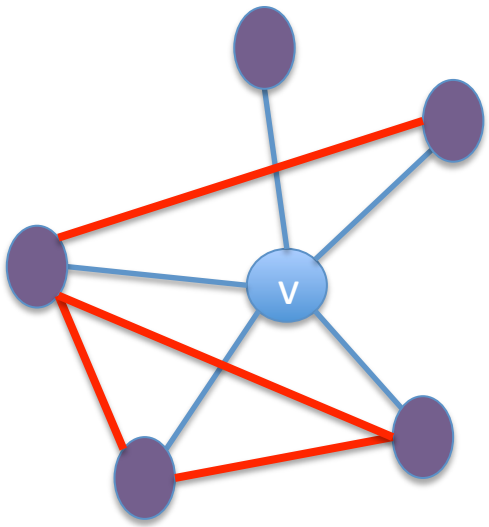
$$= 4 / {}_5C_2$$

$$= 0.4$$

クラスター係数

「自分の友達と、自分の友達は、けっこう、友達同士のことが多いよね？」

ある頂点に着目



頂点　v のクラスター係数 Cv

$$= \frac{\text{vの隣接ノード間のエッジの総数}}{\text{完全グラフの場合のエッジ数}}$$

グラフの　クラスター係数　C

Cv　の平均

レポート・成績について

- ほぼ毎回、プログラミング課題を出題する予定
  - 効率の良い計算機実験のためのツールを使ってみる
  - アルゴリズムの実装
  - ライブラリの利用・・・　など

- ３回以上、レポートのファイルとプログラムのソースを添付し、メールで提出のこと
  - E-mail: algorithm2014@edu.jar.jp
  - サブジェクト 「アルゴリズムとプログラム実践講座・レポート」
  - 学生証番号と名前は、 メールの本文にも書いてください。
  - 〆切：次の授業の直前の日曜日深夜 （講評の都合上。〆切後も受付ます）
  - プログラムやレポートは（見本として）公開することがあります。適宜、作者名やコピーライトをいれておいてください。公開不可の場合は、プログラムの冒頭にその旨、コメントをいれておいてください。
  - 質問・作問提案も歓迎 （作問については採用の場合は別途加点）
  - サンプルプログラムは「初心者向け」です。 **上級者は無視してください。**

推奨環境など

- Linux, Mac,（Windows+Cygwin）
- 仮想マシン環境（VMware, VirtualBox, Parallels)
  - 余裕があれば、いろいろな組み合わせを試して比較してみると面白いと思います

- 言語
  - 自由。ただし、一般的でない言語については、上記いずれかのOS 上にインストール可能なもの

第五回　課題

好きな問題を選んでください

その1
ランダムグラフの中の
三角形の数を数える

- 第四回の課題で作ったランダムグラフ生成器を用いてグラフを生成、生成されたグラフに、いくつ三角形が含まれているかを数え上げよ。

その２
ランダムウォークの表示

数直線の上を移動する点aを考えよう。

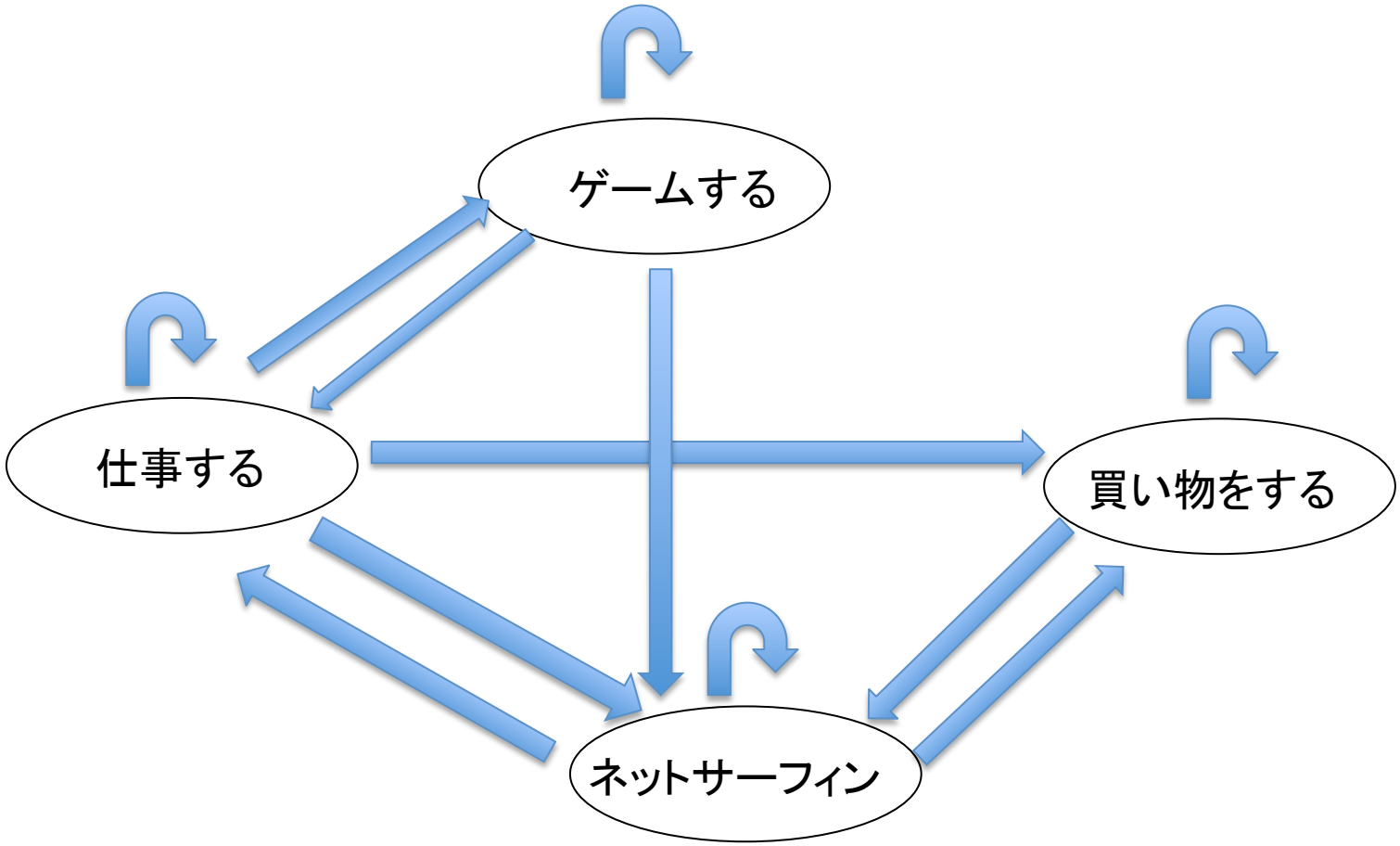
(1) 点aは　単位時間ごとに、等確率で、「＋１移動」か 「–1　移動」する。原点から始め、100単位時間後、aがどこにいたかを、適当な回数試行して、表示せよ。

(2) 点a が等確率で「＋１移動」、あるいは「静止」の場合はどうか？

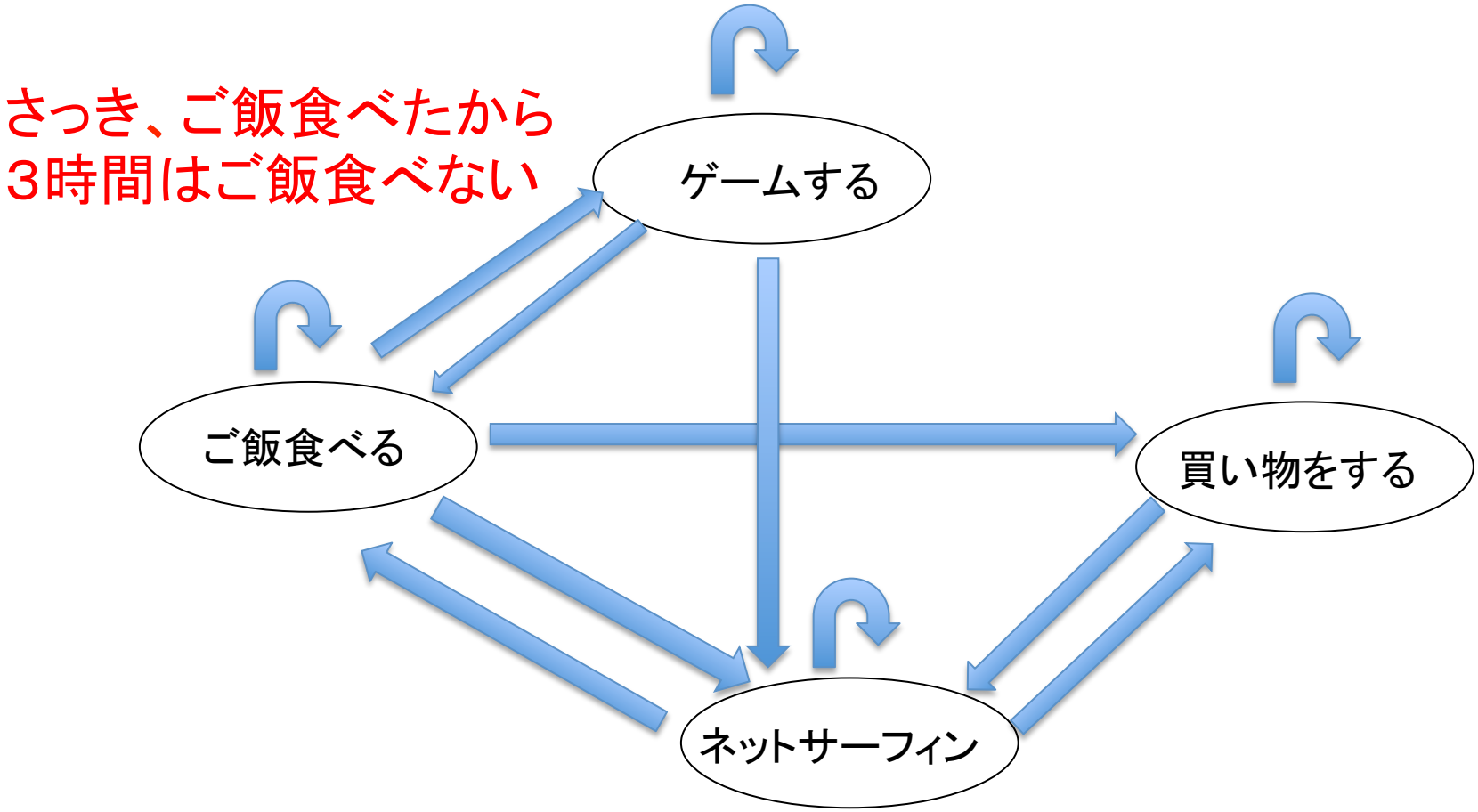
(3) 点aが　単位時間ごとに、確率　0.99 で「＋１移動」、確率0.01 で「静止」の場合はどうか？

(4) 点a が数直線ではなく、二次元のグリッド上を等確率で隣接点に移動する場合はどうか？適当な回数試行して表示せよ。また、動いた軌跡の例を表示せよ。

マルコフ過程



これは、マルコフ過程じゃない



マルコフ過程

「今、どういう状態か」 だけで、
「次の状態」が確率的に決まる。

遷移行列　P:
グラフの隣接行列の、「１」のかわりに
「そのノードに遷移する確率」をいれたもの
out-edge の確率（各列）の和は　１

状態遷移行列　P　を使うと

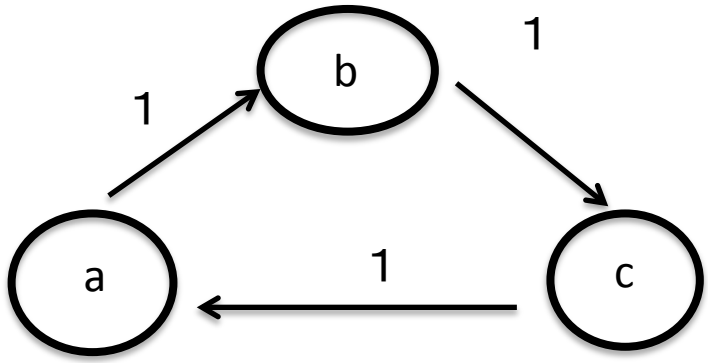
ある時刻　t　の状態を　π(t)　としたとき

$$\pi(t+1) = P\,\pi(t)$$

なら、πは 定常分布(Stationary Distribution)
（→　一つとは限らない）

周期性を持つこともある

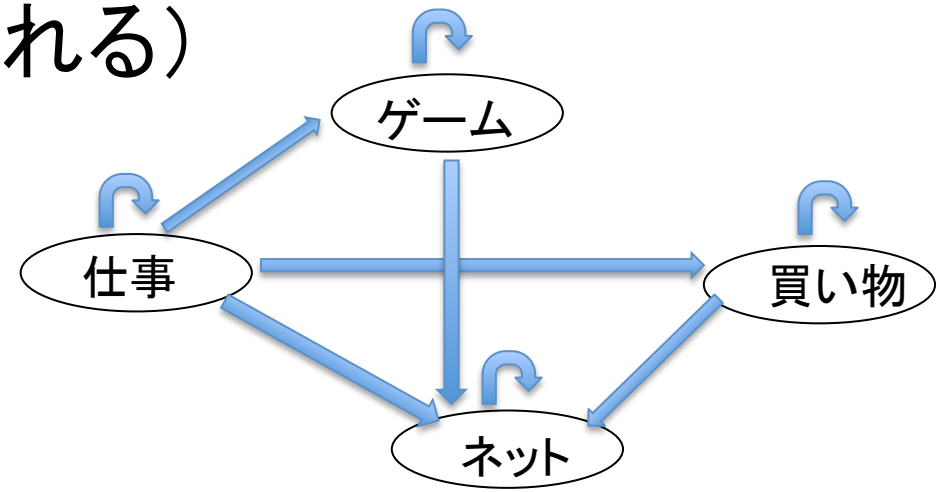
$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$



周期　３

マルコフ連鎖

- 既約 （全部、どの状態にも、たどりつける）
- 有限再帰 （もどってこれる）
- 非周期的



の場合は、**定常分布 をただ一つ持つ**

（どういう分布から始めても、その状態に落ち着く）