DDBA复习

Network

1. Links
    a. Maximum number of links

$$L_{\max} = \binom{N}{2} = \frac{N(N-1)}{2}$$

2. Density:

$$d = \frac{L}{L_{\max}}$$

3. Degree:

    In undirected network, the **degree k** of a node is the number of links a node has

    或者说, k is the number of neighbours a node has

    1. **Average degree:**

$$\langle k \rangle = \frac{\sum_i k_i}{N}$$

1. In degree: K_{in}
2. Out-degree: K_{out}
3. Total-degree: K_{tot}

2. **Strength**: (in weighted network)

    In a weighted network, a node has a **strength**:

$$S_i = \sum_j w_{ij}$$

Directed weighted network

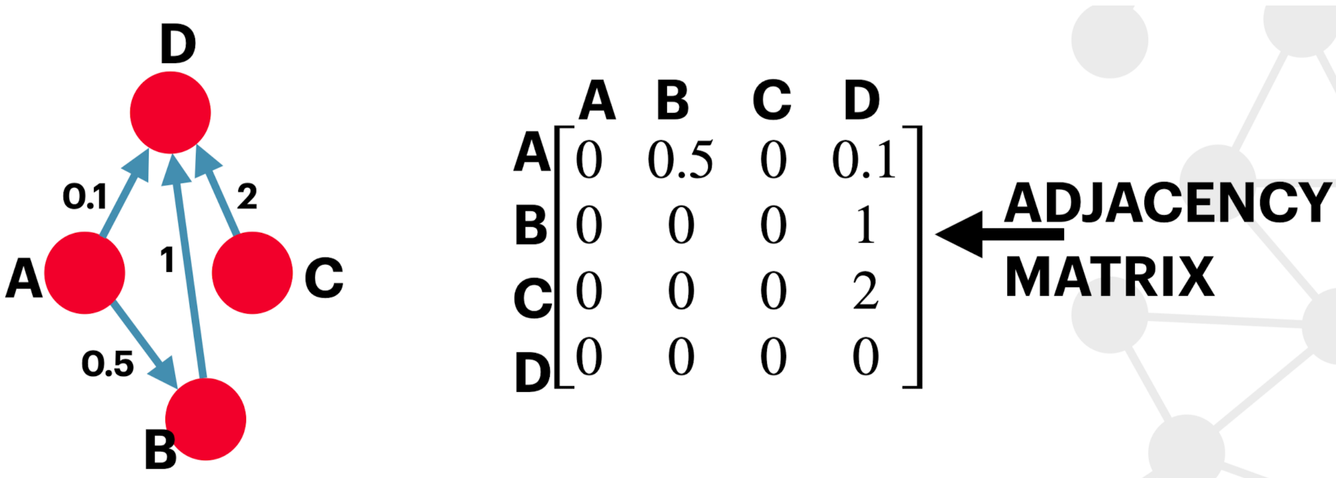
1. In-strength

$$S_i^{\text{in}} = \sum_j w_{ji}$$

2. Out-strength

$$S_i^{\text{out}} = \sum_j w_{ij}$$

3. Adjacency matrices 邻接矩阵



4. Other networks

a. Sub-network

b. Ego-network

c. Multilayer network

d. Temporal network

Theory

1. **Assortativity 同配性** 用于考察度值相近的顶点是否倾向于相互连接

   HOMOPHILY 同质性 is what causes assortativity

   一些网络没有assortative: World wide web, Ecological networks, Biological networks

   1. **Degree Correlation:**

      Nodes with **high (low) degree** connect to other nodes with **high (low) degree**.\

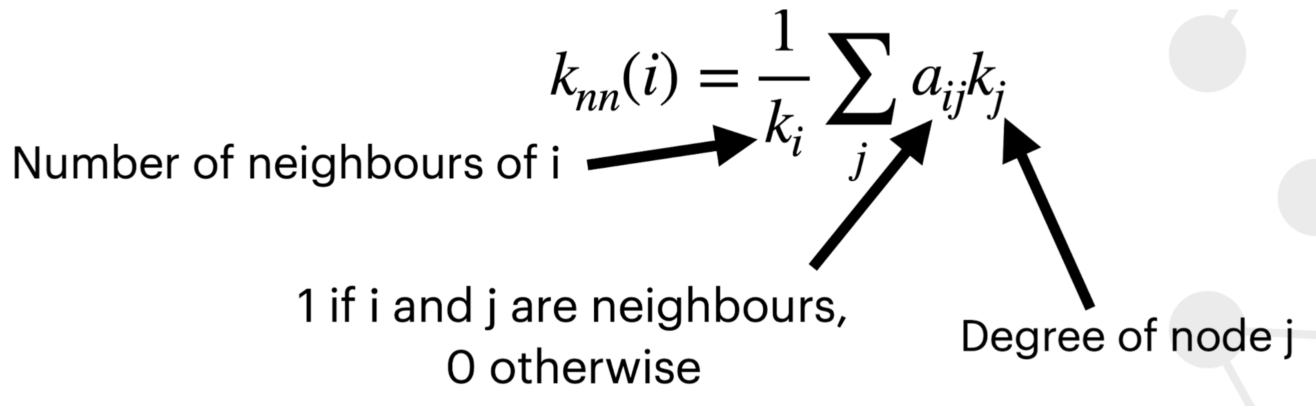
   2. Compute assortativity:

      a. **Correlation** between degrees of pairs of nodes (usually Pearson)

      b. **Average degree** of neighbours

   3. **K-nearest neighbours**

$$k_{\mathrm{nn}}(i)=\frac{1}{k_i}\sum_j a_{ij}k_j$$



   4. **K-nearest neighbours function** <$k_{\mathrm{nn}}(k)$>

      The average degree of the neighbours of nodes of degree k

2. **Paths:**

   a. **Average path - undirected network**

      所有的links 除以 最大links数

$$\langle\ell\rangle=\frac{\sum_{i,j}\ell_{ij}}{\binom{N}{2}}=\frac{2\sum_{i,j}\ell_{ij}}{N(N-1)}$$

   b. **DIAMETER - undirected network** 网络的直径（即最长的最短路径）

$$\ell_{\max}=\max_{i,j}\ell_{ij}$$

   c. An average path length is said to be short, if $\langle\ell\rangle\approx\log(N)$

3. **Connectedness 连接性**

- If there is at least a **path** between **any pair of nodes,** network is connected.
- A **component** is a **connected sub-grapgh**
- The **largest** component is called **Giant Component**
- For Directed network:
    - Weakly connected:

      if connected Only **Disregarding the direction** of links
    - Strongly connected:

      if connected Also when **considering the direction** of links

- **Clustering coefficient 聚类系数 节点倾向于聚类程度的度量**

  C(i) of node i is the fraction of pairs of the neighbours of i that are connected to each other. **节点 i 的 C(i) 是 i 的邻居中相互连接的成对邻居的比例。**

$$C(i) = \frac{\tau(i)}{\tau_{\max}(i)} = \frac{2\tau(i)}{k_i(k_i - 1)}$$

- Trees

  Maximum (minimum) spanning tree:

  For each node only keep the connection with maximum (minimum) weight

4. **Heterogeneity 异质性**
    - In real-world networks the importance of nodes is heterogeneous
    - Importance is often measured with **centrality**
        - **Closeness centrality 接近中心度**

          节点与其他节点之间最短路径长度之和的倒数

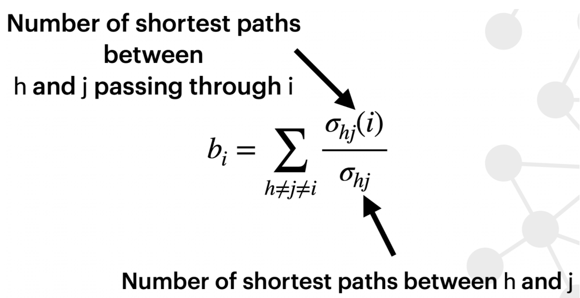
$$g_i = \frac{1}{\sum_{i \neq j} \ell_{ij}}$$

$$\tilde{g}_i = (N-1)g_i = (N-1)\frac{1}{\sum_{i \neq j} \ell_{ij}} = \frac{1}{\sum_{i \neq j} \frac{\ell_{ij}}{(N-1)}}$$

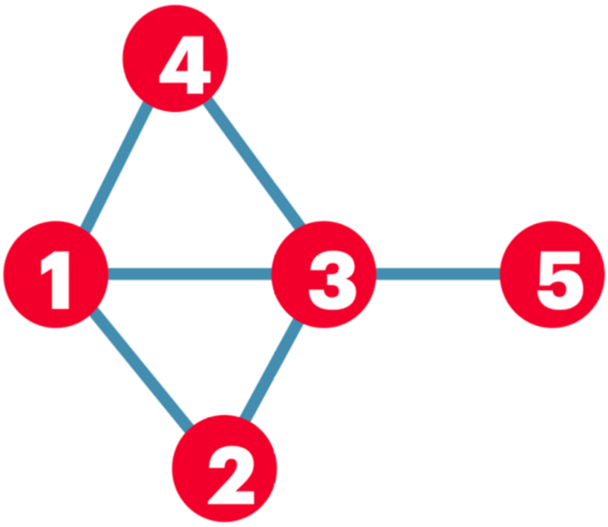
        - **Betweenness centrality 介度中心性**

          How many shortest paths pass through a node

$$b_i = \sum_{h \neq j \neq i} \frac{\sigma_{hj}(i)}{\sigma_{hj}}$$



            - EX:



$$k_3 = 4$$

$$g_3 = \frac{1}{4}$$

$$b_3 = 3.5$$

$$g_3 = \frac{1}{\ell_{1,3} + \ell_{2,3} + \ell_{4,3} + \ell_{5,3}} = \frac{1}{1+1+1+1} = \frac{1}{4}$$

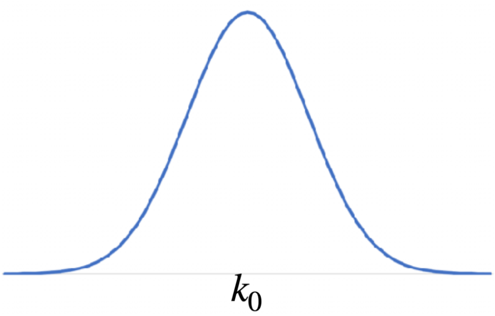
~~1,2~~ ~~1,3~~ ~~1,4~~ 1,5
~~2,3~~ 2,4 2,5
~~3,4~~ ~~3,5~~
4,5

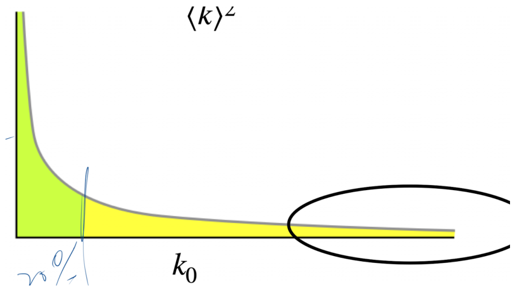
$$b_3 = \frac{\sigma_{1,5}(3)}{\sigma_{1,5}} + \frac{\sigma_{2,4}(3)}{\sigma_{2,4}} + \frac{\sigma_{2,5}(3)}{\sigma_{2,5}} + \frac{\sigma_{4,5}(3)}{\sigma_{4,5}} = 3.5$$

- **Measure Heterogeneity 测量异质性**

$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle^2}$$

  - $\langle k \rangle$ 表示网络中节点度的平均值。
  - $\langle k^2 \rangle$ 表示网络中所有节点度的平方的平均值。





- If not heterogeneous $\langle k^2 \rangle \approx \langle k \rangle^2 \approx k_0^2$ $\quad k \approx 1$
- if heterogeneous $\langle k^2 \rangle \gg \langle k \rangle^2$ $\quad k \gg 1$

Build Network

Random Network

Gilbert and **Erdos and Rényi,** both called **Erdos-Rényi networks(ER networks)**

- **Gilbert model**
  - **Parameters**: **n** number of nodes and a probability **p**
  - **Algorithm:**
    - For each pair of nodes i,j:
    - Generate a random number r between **0** and **1** (**uniformly distributed**)
    - If r<p make a link between i,j
- **Erdos-Renyi model**
  - **Parameters**: **n** number of nodes and **L** number of links
  - **Algorithm:**
    - Build all possible networks with n nodes and **L** links Pick one randomly
  - **Remarks:**
    - For large values of **L**, these two formulations are (almost) equivalent

- Density:

$$\langle L\rangle = p\binom{N}{2} = p\frac{N(N-1)}{2}$$

$$\langle k\rangle = \frac{2\langle L\rangle}{N} = p(N-1)$$

$$\langle d\rangle = \frac{\langle L\rangle}{L_{\max}} = \frac{p\frac{N(N-1)}{2}}{L_{\max}} = \frac{p\frac{N(N-1)}{2}}{\frac{N(N-1)}{2}} = p$$

  - $\langle L\rangle$是网络中期望的链接数，通过连接概率p乘以所有可能的链接数来计算。
  - $\langle k\rangle$ 是网络中节点的平均度数，等于两倍的$\langle L\rangle$除以节点数 N。
  - $\langle d\rangle$ 是网络的平均路径长度，等于 $\langle L\rangle$ 除以最大路径长度 $Lmax$。
- Degree Distribution:
  - **Binomial distribution 二项分布**

$$P(k) = \binom{N-1}{k}p^k(1-p)^{N-1-k}$$

  在这个模型中，网络的每对节点之间以概率 p 相连。对于网络中的一个节点：

    - $P(k)$ 表示一个节点恰好有 k 个连接的概率
    - $N-1$ 是除了节点本身之外的网络中其他节点的总数。
    - $\binom{N-1}{k}$表示从 $N-1$个节点中选择 k 个节点的方式数。
    - p^k是 k 个节点都被选中的概率。
    - $(1-p)^{N-1-k}$是剩下的 $N-1-k$ 个节点都不被选中的概率。

- Short Paths
  - **Assumption:**

    在所有节点的度都是 k 的假设下，可以到达的节点数量随路径长度 ℓ 的增加而呈指数增长。

Path length	**Nodes reached**
ℓ= 1	*k*

ℓ= 2	k(k− 1)
ℓ= 3	k(k− 1)2
If k large	$k(k-1)^{\ell-1} \approx k^{\ell}$

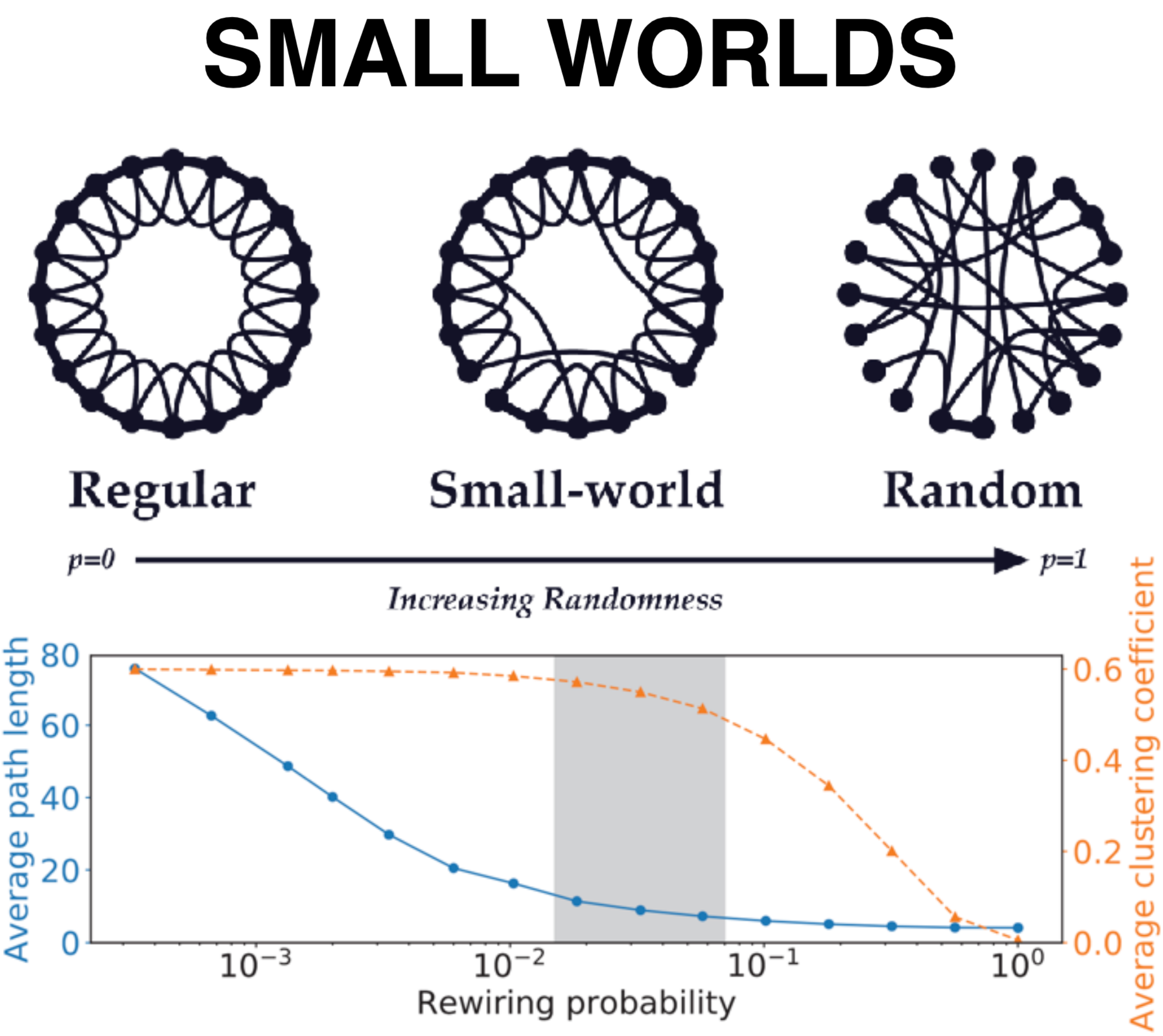
最大路径长度ℓ_{max}的计算方法，即使在节点数量很大的网络中，也能保持相对较小的路径长度。Diameter grows with logarithm of nodes, so distances are fairly small despite the network size.

$$k^{\ell_{max}} = N$$

$$\ell_{\max} = \log_k N = \frac{\log N}{\log k}$$

Small World Network

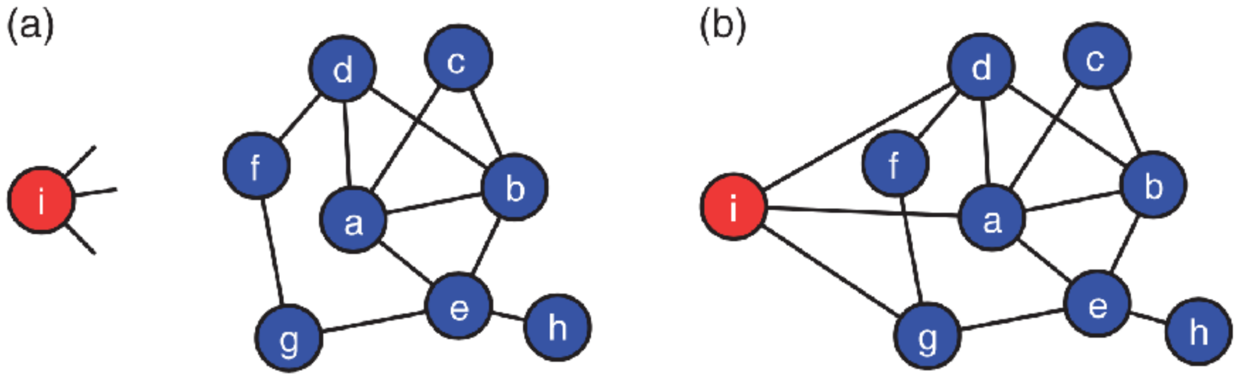
- “小世界”网络是一种数学模型，它描述了尽管网络可能很大，但网络中大多数节点之间可以通过相对较少的步骤相互到达。这种类型的网络具有两个主要特征：高聚类系数和短平均路径长度。高聚类系数意味着一个节点的邻居节点很可能互相连接；短平均路径长度意味着从网络中任意一个节点到任意另一个节点的步数通常很少。这种网络的结构在社交网络、神经网络以及某些类型的技术网络中普遍存在。
- Proposed by **WATTS AND STROGATZ** (sometimes called WS networks), Solves the problem of low **clustering coefficients**



SCALE-FREE NETWORKS

- Proposed by REKA **ALBERT AND** LASZLO **BARABASI (**BA NETWORK**)**
- Based on the idea of **PREFERENTIAL ATTACHMENT 优先附着**
- Introduces **dynamic** generation of networks and **explains hubs**

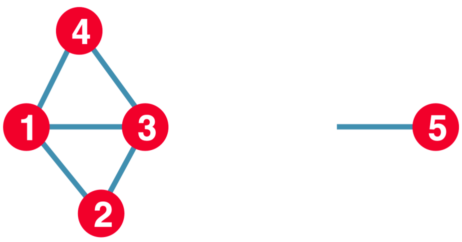
- **无标度网络**是一种其度分布遵循幂律分布的网络模型，至少在渐近意义上是这样。这意味着少数节点（称为枢纽节点）具有很多连接（高度数），而大多数节点只有少数连接。这种网络在许多现实世界的系统中都有发现，例如互联网、社交网络和生物系统。"无标度"这个概念来源于这样一个事实：这些网络在任何规模上看起来都相似，幂律分布没有与之关联的典型规模。
- **PREFERENTIAL ATTACHMENT 偏好连接**



NEW NODES MAKES SOME CONNECTIONS

CONNECTIONS ARE MADE BASED ON A GIVEN RULE (OR SET OF RULES)

- **BARABASI-ALBERT MODEL**



$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

Probability a new node makes a connection to node i

$\Pi(2) = \frac{2}{2+2+3+3} = \frac{1}{5}$

For nodes 2 and 4

$\Pi(3) = \frac{3}{2+2+3+3} = \frac{3}{10}$

For nodes 1 and 3

  - **Parameters**:

Number of initial nodes	m_o
Number of links of new nodes	$m \leq m_0$
Number of total nodes at the end of the process	N

  - **Algorithm:**

For each new node i , with i from m_o to: N
    Compute the probability of every node to get attached to i based on degree
    Choose m nodes randomly based on this probability
    form link between the new node and the chosen nodes

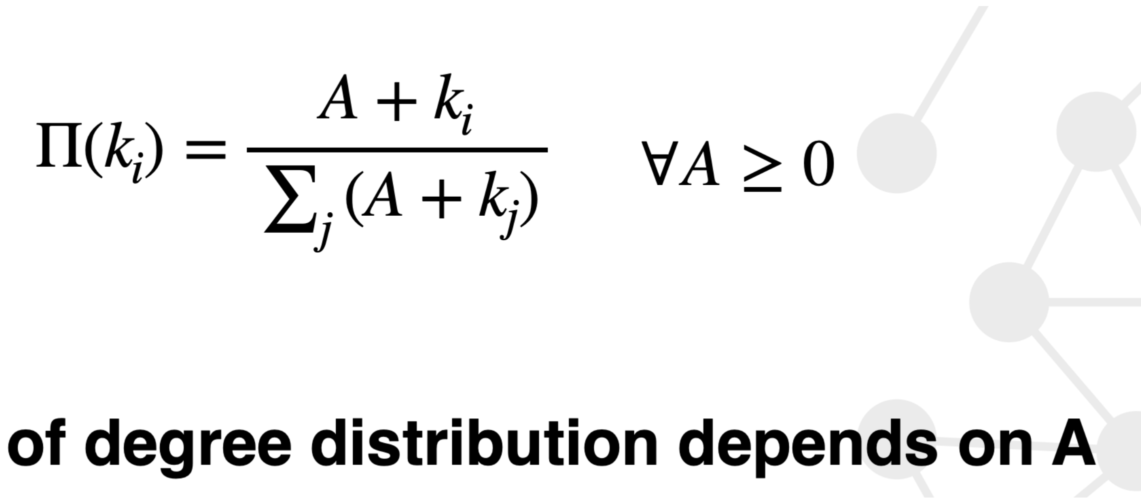
  - **Metrics:**
    - **Degree distribution 度分布**: $P(k) \approx k^{-3}$
    - **Clustering coefficient**: $C(k) \approx k^{-1}$

- **Diameter:** $\ell(k) \approx \frac{ln(N)}{lnln(N)}$
  - **Limitation:**
    - Fixed degree **distribution slope** 固定度分布坡度
    - The **oldest nodes** become hubs 最老的节点成为枢纽
    - **Clustering lower** than real-world networks 聚类能力低于实际网络
    - **Nodes/Links** cannot be removed
    - Single **connected component**
    - Limiting **initial conditions**
  - **Non-linear Preferential Attachment**

$$\Pi_\alpha(k_i) = \frac{k_i^\alpha}{\sum_j k_j^\alpha}$$

α = 1	Barabási-albert model
α < 1	No hubs
α > 1	Fewer, larger hubs

- **Attractiveness model**



- **Fitness model**
  - is an extension of the **preferential attachment model 偏好连接**
  - 认为形成新链接的概率不仅取决于每个节点的度，还取决于称为 "适配度 "的节点特定属性。
  - 每个节点都有一个适合度值，代表其吸引链接的能力。在这一模型中，"适应度 "值越高的节点越有可能获得新的连接，从而使最 "适应 "的节点成为网络枢纽，这能更好地反映现实世界网络的复杂性。

$$\Pi(k_i) = \frac{\eta_i k_i}{\sum_j (\eta_j k_j)}$$

Fittest, not oldest nodes become hubs

We take η from a distribution $\rho(\eta)$

If$\rho(\eta)$ has a finite support (finite maximum value) there are hubs. Otherwise, there will likely be a single massive hub.

- **Configuration model**
  - 每个节点被赋予一个固定的度数，这个度数是从一个给定的度分布中选取的。
  - 然后通过随机连接这些度数来构造网络，直到所有节点达到其指定的度数。
  - 网络是随机的，但度分布是固定的
  - Takes a **degree sequence** as input
  - Links are **rewired** randomly, degree sequence preserved
  - Very important for **benchmarking**!

Community

Definitions:

- **Internal and External degree:**
  - The number of neighbours inside and outside the community

$$k_i = k_i^{int} + k_i^{ext}$$

    - **i is called internal node of community c if** $k_i^{ext} = 0$ **and** $k_i^{int} > 0$
    - **i is called boundary node of community c if** k_i^{ext} **> 0 and** $k_i^{int} > 0$
- **Number of internal links:**
  - The number of links between nodes within the community
- **Community degree:**
  - The sum of degree of all the nodes in the community

$$k_C = \sum_{i \in C} k_i$$

- **Internal link density:**
  - Density that considers only links between members of the community

$$\delta_C^{\text{int}} = \frac{L_C}{L_{C_{\max}}} = \frac{2L_C}{N_C(N_C - 1)}$$

- **Strong Community**
  - 如果社区中的每个节点 i 的内部度 k_i^{int} 都大于其外部度 k_i^{ext}，则该社区被认为是强社区。

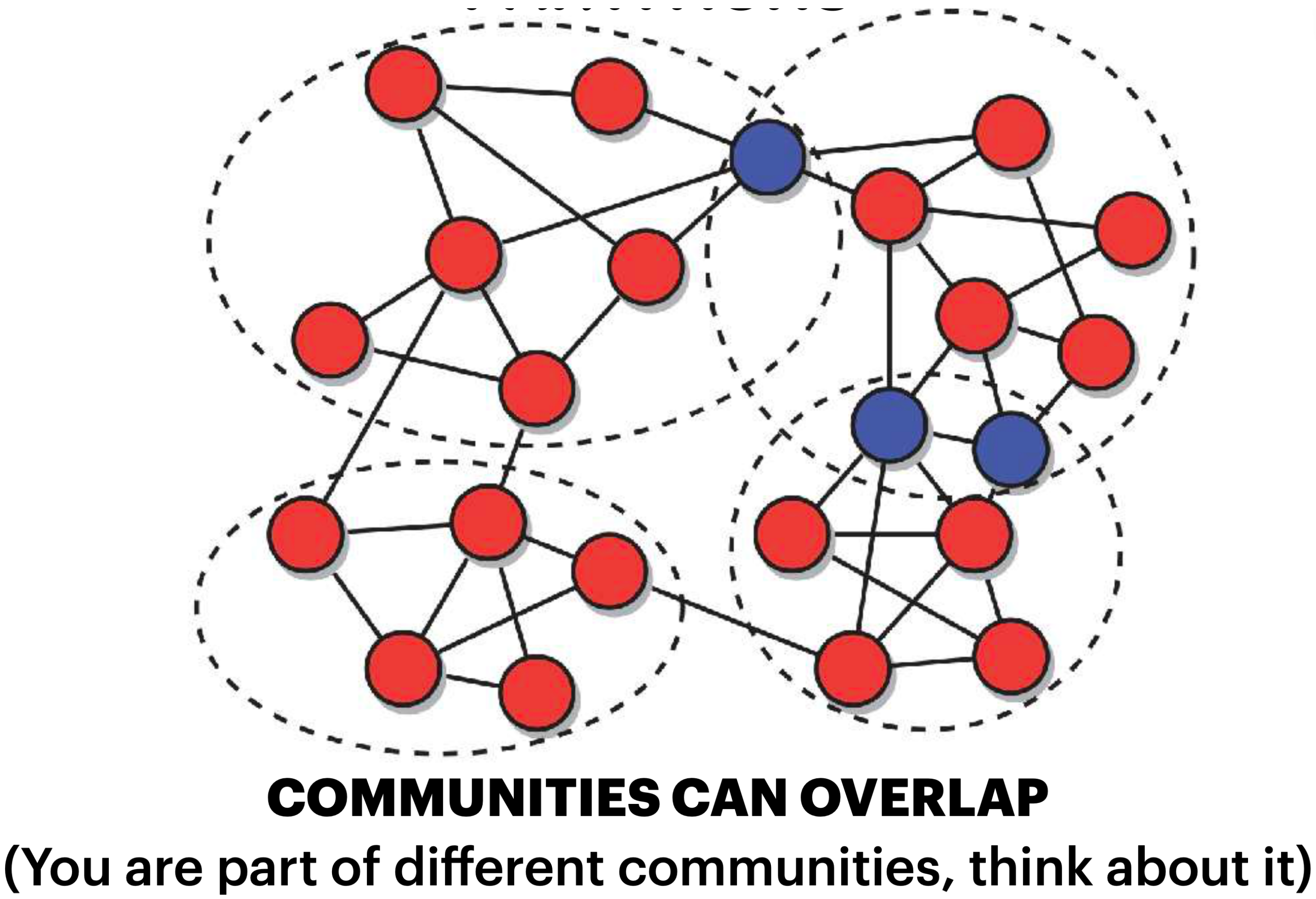
$$\forall i \in C : k_i^{\text{int}} > k_i^{\text{ext}}$$

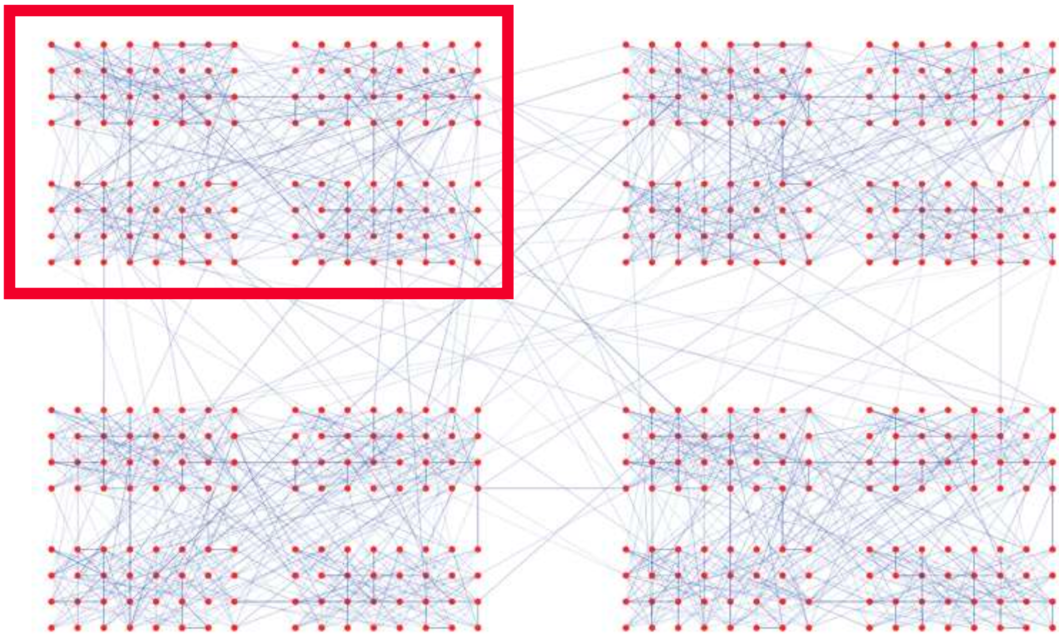
- **Weak Community**
  - 如果社区中所有节点的内部度之和大于外部度之和，即 $\sum_{i \in C} k_i^{\text{int}} > \sum_{i \in C} k_i^{\text{ext}}$，则该社区被认为是弱社区。

$$\sum_{i \in C} k_i^{\text{int}} > \sum_{i \in C} k_i^{\text{ext}}$$

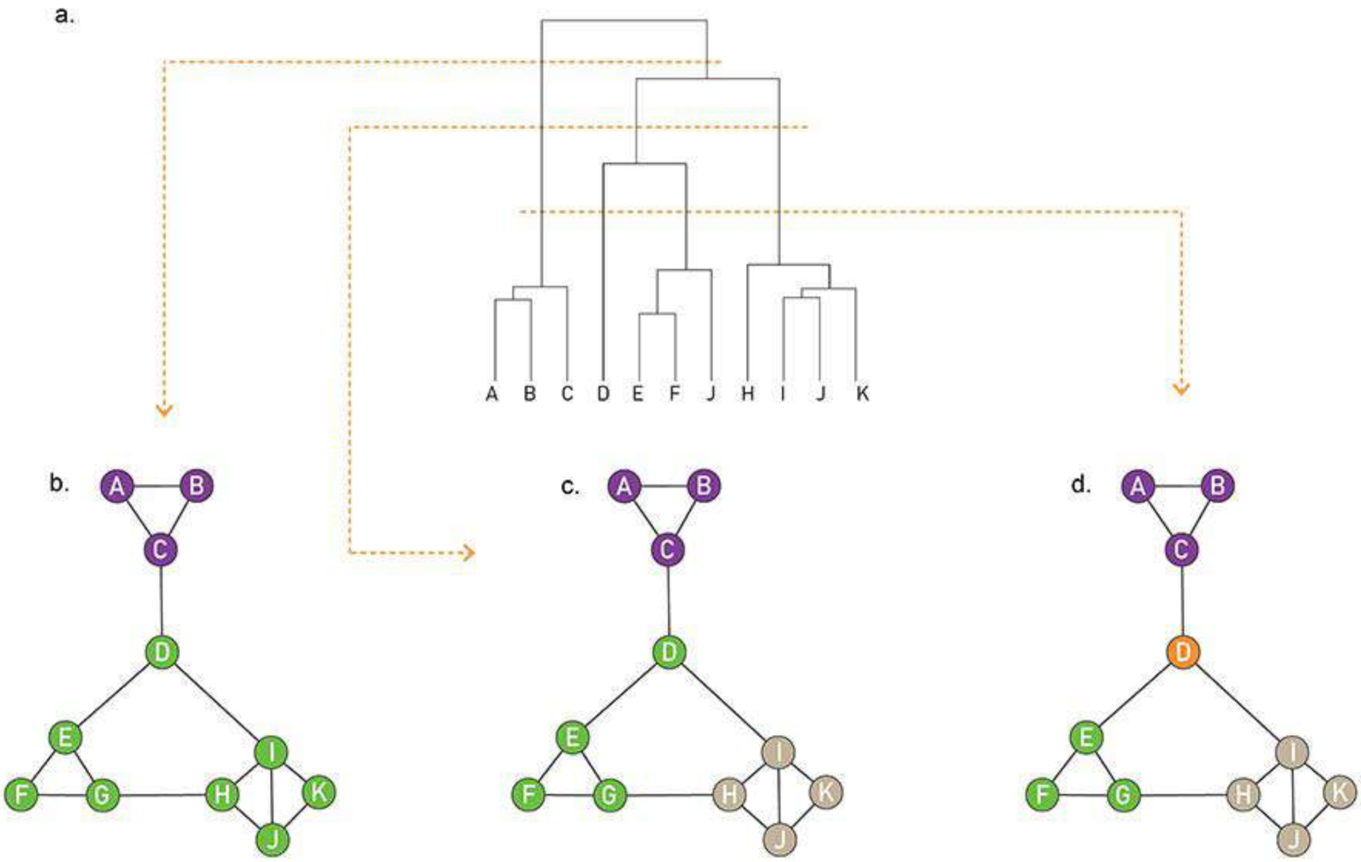
Partitions

- Partitioning is the division of a network into communities





COMMUNITIES CAN BE HIERARCHICAL
(There might be communities within communities)



Dendrogram

Community Detection

- **Approaches:**
  - Bridge removal
  - Modularty maximisation
  - Label propagation 标签传播
  - Stochastic block modelling 随机区块建模
- **Bridge removal:**
  - A bridge is a link whose removal breaks the network into two parts
  - **Girvan-Newman algorithm**
    1. compute link **betweenness** for all the links
    2. **remove** the link with highest betweenness
    3. **repeat** 1 and 1 until you have no links left

    COMPUTING LINK BETWEENNESS FOR LARGE NETWORKS IS IMPOSSIBLE
- **Modularty maximisation:**
  - **Main idea:** Calculate how good a community is vs Random Baseline

    Difference between links in c and expected links in c with configuration model

$$Q = \frac{1}{L}\sum_{C}\left(L_C - \frac{k_C^2}{4L}\right)$$

    - $\frac{k_C}{2L}$ Is the probability of randomly choosing **one stub** in the community
    - $(\frac{k_C}{2L})^2$ Is the probability of randomly choosing **two stubs** in the community
    - There are **L** links in the network
    - Each link joins two stubs from community c with probability $(\frac{k_C}{2L})^2$
    - Then, the **expected number** of links in the community is $L(\frac{k_C}{2L})^2 = \frac{k_C^2}{4L}$

    公式中的 $\frac{k_C^2}{4L}$ 是在配置模型下，社区 C 内部链接数的期望值。模块性 Q 是所有社区 C 内部实际链接数与期望链接数之差的总和，然后除以 L。当 Q 达到最大值时，社区划分被认为是最优的。
  - **Directed**

$$Q_d = \frac{1}{L}\sum_{C}\left(L_C - \frac{k_C^{\text{in}} k_C^{\text{out}}}{L}\right)$$

  - **Weighted**

$$Q_w = \frac{1}{W}\sum_{C}\left(W_C - \frac{{s_C}^2}{4W}\right)$$

  - **Weighted and Directed**

$$Q_{dw} = \frac{1}{W}\sum_{C}\left(W_C - \frac{s_C^{\text{in}} s_C^{\text{out}}}{W}\right)$$

  - Most famous algorithms: **Louvain, Leiden**

1. start with no communities. Every nodes is moved to a community so that Q Is maximised. Repeat until no modularity gain is possible 开始时没有社区。每个节点都被移到一个群落中，从而使 Q 最大化。重复上述步骤，直到无法获得模块化增益为止
2. the network becomes a weighted super-network, in which nodes are the communities of the original network, and weights are the number of links between communities (this includes self-loops) 网络成为加权超级网络，其中的节点是原始网络中的社区，权重是社区之间的链接数（包括自循环）。

- **Problems:**
    - **Comparison:** On average Larger networks have larger modularity 平均而言，网络越大，模块化程度越高
    - **Uncertainty:** this approach can find positive modularity for random networks 这种方法可以找到随机网络的正模块性
    - **Resolution:** cannot find communities whose degree is smaller than $\sqrt{2L}$

- **Label propagation:**
    - **Process**:
        1. Start with Singletons 从单子开始
        2. One by one, with random order, nodes take the “Label” (IE Community membership) of the **major of their neighbours** 节点按随机顺序一个接一个地获取其邻居主要节点的 "标签"（即社区成员资格）。
        3. Repeat this unit the **Partition is stable** (IE there are no possible changes) 重复该单元，分区保持稳定（即没有可能的变化）
    - **Issues:**

      Different runs find different communities needs to be run multiple times
    - **Strengths:**
        - Very fast
        - If some memberships are known, they can be used to initialise the network
- **Stochastic block modelling:** 随机区块建模
    - Generative algorithm
    - Generates communities with given probabilities, chooses the most likely
    - Can perform community detection on a lot of different network types
        - ex: IF $\forall r, p_{rr} = 0$, This represents multipartite networks 多部分网络
    - And can discover **more than just communities:**

∀r, s $p_{rr} > p_{rs}$	**Classic communities**
$p_{rr} < p_{rs}$	**Disassortative structure 非同质性结构**
∀r $p_{rr} = 0$	**Multipartite network 多部分网络**
∀r,s $p_{rr} = p_{rs} = p$	**Random network**

    - **Limits:**
        - Need prior knowledge of number of communities 需要事先了解社区数量
    - **Strength:**
        - EVERYTHING ELSE

Simple models, complex models

- **Contagion: 传染**
    - Simple contagion: Agents can be “infected” directly
    - Complex contagion: Non-linear or repeated interactions for agents to be “infected”
- **Independent cascade: 独立级联模型**

- **Basic:**
  - 是一个用于模拟信息、影响或者疾病在网络中传播的模型。在这个模型中，每个已经被影响的节点有单独的机会去影响其邻居节点。如果节点A影响了节点B，那么节点B在下一轮中有机会去影响其他的节点，依此类推。这个过程持续进行，直到没有更多的节点可以被影响.
  - **Irreversible** states
  - Easy to **compute**
  - Easy to **extend** to influence maximisation
- **Good for:**
  - information and rumor propagation 信息和谣言传播

- **Linear threshold model:**
  - **Basic:**
    - 是一种用于社交网络中的信息或行为传播模型。在此模型中，每个节点根据其邻居的影响程度和预设的阈值来决定是否改变状态。具体来说，每个节点都有一个阈值，当其接收到的总影响超过这个阈值时，它就会被激活。这个模型通常用来模拟社会行为的变化，如人们如何决定采纳一个新的技术或者观念。
    - **Irreversible** states
    - Easy to **compute**
    - Toy model for **complex contagion** 复杂传染的玩具模型
  - **Good for:**
    - Group decisions, Majority voting, etc.

- **Voter model 选民模式**
  - **Basic**:
    - 是一个简单的数学模型，用于模拟意见形成和传播的动态过程。在这个模型中，网络中的每个节点代表一个"投票者"，每个投票者持有一个意见。在每个时间步骤中，随机选取一个节点，它将从其邻居中随机选择一个，然后采纳这个邻居的意见。这个过程不断重复，模拟了在社交互动中意见如何逐渐统一。投票者模型常用于研究政治动态、文化传播以及共识形成等领域。
    - **Flipping** states 翻转状态
    - **Simple**, analytical solution for equilibrium 简单的平衡分析法
    - **Fails** when opinions are fine-grained 意见细化时失败
  - **Properties:**
    - "Always" reach consensus
    - Analytical formula for equilibrium
    - Many possible extensions
  - **Process:**
    1. Pick a random node i
    2. pick a random neighbour j of node i
    3. change opinion of i to opinion of j
  - **Good for:**

    Anything with binary states that can be reverted 任何二进制状态都可以还原

- **Bounded confidence models 有界信任模型**
  - **Basic:**

- 是社会动力学领域中的一类模型，它们模拟个体如何仅在意见相近的情况下才相互影响。在这些模型中，每个个体有一个意见或信念的数值，而个体间的互动仅当他们的意见差异小于某个阈值（称为信心界限）时才发生。这意味着个体只考虑与自己意见相近的邻居的影响。这些模型通常被用来研究意见分化、群体极化或共识形成等现象。
- **Continuous** opinions 持续的意见
- Usually between **0 and 1** but not necessarily
- **Easy to implement** with agent-based models

- **Deffuant model**
  - 是一个有界信任模型，用于模拟个体意见的动态变化和演化。在这个模型中，个体被赋予一个连续的意见值，当两个个体的意见差异在某个阈值（信任阈值）以内时，它们会相遇并更新彼此的意见。通常，这种更新会导致他们的意见值向对方靠拢，这个过程可以看作是一种意见的平均或者妥协。Deffuant模型经常被用来研究如何从初始的意见多样性中达成某种形式的共识，或者网络中意见极化的形成过程。
  - Equations

    这个模型中，两个意见代理（或者个体）i 和 j 只有在它们的意见差距小于某个阈值 ϵ 时才会互相影响。互动后，它们的意见会更新为彼此意见的平均值。

Opinion agent i

Equations

Convergence parameter

$$\begin{cases} o_i(t+1) = o_i(t) + \mu(o_j(t) - o_i(t)) \\ o_j(t+1) = o_j(t) + \mu(o_i(t) - o_j(t)) \end{cases}$$

Only If $|o_i(t) - o_j(t)| < \epsilon$

  - **Good for:**
    - when you need continuous opinions

Network structure comparison

- Scale-free networks are **more robust 更坚固**
  - Most nodes have **low degrees**
  - **Hubs are highly connected and central**

Targeted removal

- **Robustness** of different networks
- **Targeting** strategies
- **Financial networks** and systemic risk

Agent-based model

A tale of rationality

- Everybody is **perfectly rational**

- We **always** make the **best** decision
- We **always know how** to make the **best** decision

The efficient market hypothesis (EMH) 有效市场假说

- Everybody is **perfectly rational**
- Price reflects all **available information**
- **Forms:** Weak, semi-strong, strong

Black-Scholes模型假设

- Returns (price changes) are gaussian
- Market price is perfectly random 资产价格的变动是随机的无法预测
- Everything can be described by one equation

ABM

- Basic:
  - Objects **interact** with each other and/or with the environment
  - Objects are **autonomous 有自主权的** (bottom-up approach)
  - The outcome of their interactions is numerically **computed**
- **Characteristics 特性**:
  - Heterogeneity 异质性
  - Explicit space 明确空间
  - Local interactions 本地互动
  - Scalability 可扩展
  - Non-equilibrium dynamics 非平衡动力学
- When to use
  - Model complex systems
  - Get quick intuition of the dynamics of the system
- Why use abm
  - **Numerical computation of analytical models** 分析模型的数值计算
    - Model is **not** analytically **soluble** for some variable对于某些变量模型无法分析解决
    - **empirical** distribution of variable needs to be compared with **theoretical** 需要将变量的经验分布与理论分布进行比较
    - **Out of equilibrium** solutions not possible analytically 无法分析非平衡解
  - **Testing robustness of analytical models** 测试分析模型的稳健性
    - Changing assumptions or values often leads to no solution or intractability 改变假设或价值往往导致无解或难以解决
  - **Stand-alone simulation models** 独立模拟模型
    - Substitute, not complementary, to mathematical analysis 取代而非补充数学分析
    - Problems are analytically intractable 问题在分析上难以解决
    - Analytical solution bears no advantage 分析解决方案没有优势
- Replace the (crazy) assumptions of neoclassical economics 新古典经济学:
  - Agents are fully rational with unbounded computational skills 代理是完全理性的，拥有无限制的计算能力
  - Equilibrium defined by re and no-arbitrage condition 由再套利和无套利条件确定的均衡
  - Interactions and heterogeneity only add noise 相互作用和异质性只会增加噪音
- Discrete event simulations (DeS)

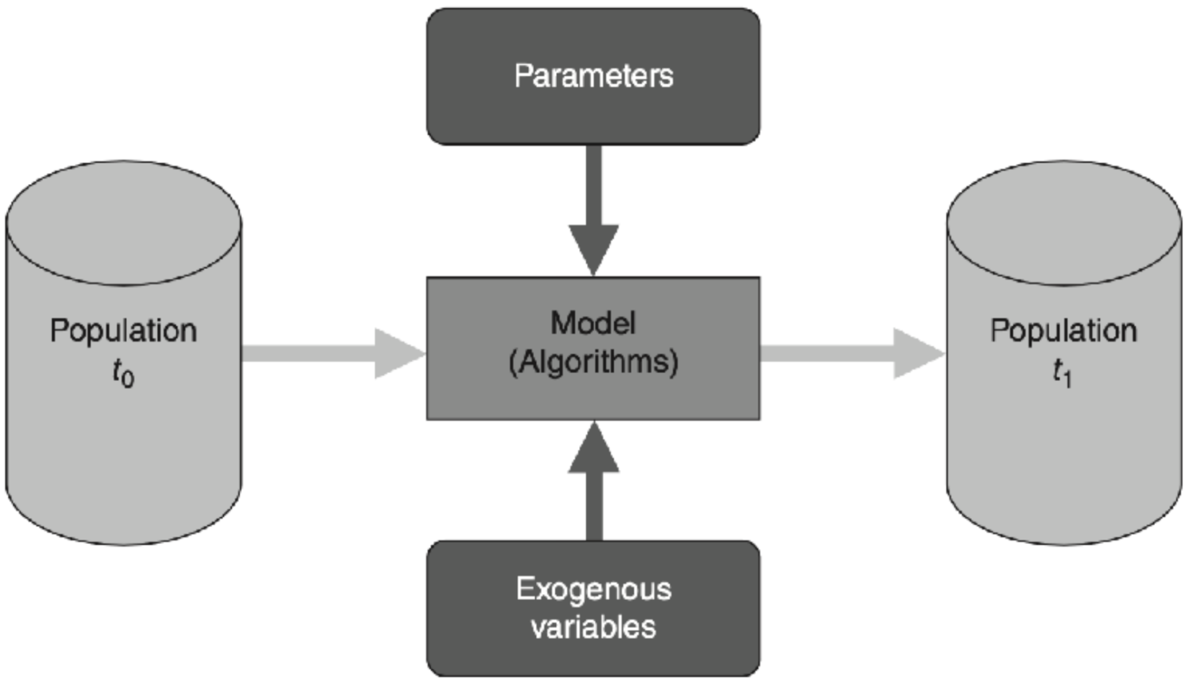
- Discrete vs continuous time

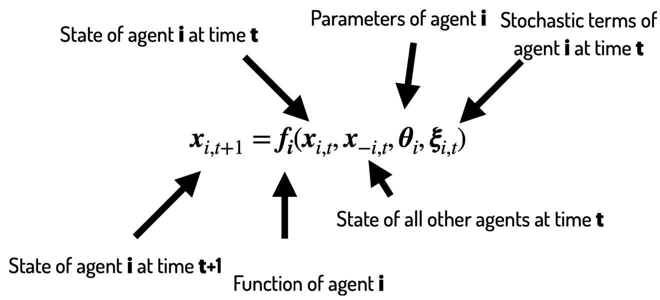
Discrete event simulations (DeS)	Continuous simulations / system dynamics (sd)
Given state at given time	Stock/flows/delays
Move between states between t and t+1	Differential equations

- ABM vs DES

Discrete-event simulations	Agent-based models
Process oriented （top-down modelling approach）;the focus is on modelling the system in detail, not the entities. 面向过程（自上而下的建模方法）；重点是详细建模系统，而不是实体。	Individual based （bottom-up modelling approach）; the focus is on modelling the entities and interactions between them. 基于个体的（自下而上的建模方法）；重点是对实体和实体之间的相互作用进行建模。
One thread of control （centralised）. 单线控制（集中式）。	Each agent has its own thread of control （decentralised）. 每个代理都有自己的控制线（分散式）。
Passive entities, i.e., something is done to the entities while they move through the system; intelligence （e.g-, decision-making） is modelled as part in the system. 被动实体，即实体在系统中移动时，系统会对实体做一些事情；智能（如决策）被模拟为系统的一部分。被模拟为系统的一部分。	Active entities, i.e., the entities themselves, can take on the initiative to do something; intelligence is represented within each individual entity. 积极的实体，即实体本身，可以主动去做一些事情；智慧体现在每一个单独的实体中。
Queues are a key element. 队列是一个关键因素。	No concept of queues. 没有排队的概念。
Flow of entities through a system; macro behaviour is modelled. 实体在系统中的流动；宏观行为建模。	No concept of flows; macro behaviour is not modelled, it emerges from the micro decisions of the individual agents. 没有流动的概念；宏观行为不是模拟出来的，而是从个体行为者的微观决策中产生的。
Input distributions are often based on collect/measured （objective） data. 输入分布通常基于收集/测量的（目标）数据。数据。	Input distributions are often based on theories or subjective data. 输入分布通常基于理论或主观数据。

- Structure of ABMs

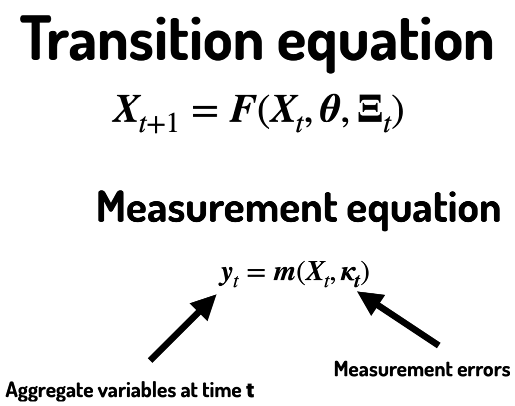




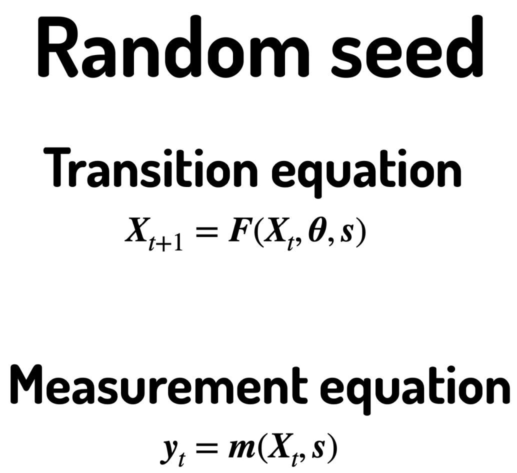
- Transition equation

$$X_{t+1} = F(X_t, \theta, \Sigma_t)$$

- State-space representation 状态空间表示法



时间 t 的总体变量



Modeling agents behaviour

Agent design

1. **Nature** of agents
2. **List of variables describing their state**
3. List of **actions** the agents can perform
4. Structure of their interaction with other agents

Experiment

1. Definition of output variables of interest 相关产出变量的定义

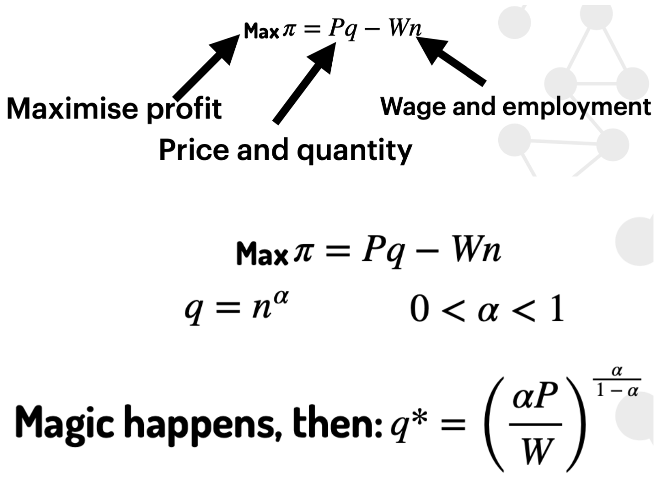
2. Appropriate experimental design 适当的实验设计
3. Analysis of equilibria 平衡分析
4. Sensitivity analysis 敏感性分析

Certainty (full rationality)

- Objective utility function 目标效用函数
- Constraints 约束
- Perfect information
- Perfect cognitive capabilities

  - Agents are fully informed about the environment
  - Agents have unbounded time and computational power
  - Agents are consistent

ex: production problem



Uncertainly

- Some variables may be **unknown** or not computable
- **Measurable or tractable uncertainty (risk)**
  - Agents know all possible states
  - Probability distribution
- **Untractable uncertainty**
  - True uncertainty
  - Don't know the states or can't compute probability

Risk neutrality

- Agents are risk neutral if in presence of measurable uncertainty they maximise the expected value of the uncertain payoff 如果存在可衡量的不确定性，代理人会最大限度地提高不确定收益的预期值，那么代理人就是风险中性的
- Agents can still make optimal choice based on available information 代理仍可根据现有信息做出最佳选择
- Agents form expectations 代理形成期望
- Expected selling price:

$$E(p) = p_h P_h + p_l P_l$$

  - Selling price in each state P_h and P_l

- Probability of each state p_h and p_l

ex: production problem

$$\text{Max}\ \pi = E(P)q - Wn$$

$$q = n^{\alpha} \qquad 0 < \alpha < 1$$

$$\text{Magic happens, then:}\ q^* = \left(\frac{\alpha E(P)}{W}\right)^{\frac{\alpha}{1-\alpha}}$$

Rational expectation models 理性预期模型

Clearly not true

Perfect **information** and **computing** capabilities
Agents will learn the **"true model"**
Representative agent

What is the true model?

Are we all the same?

Heterogeneous beliefs 各种各样的

No. Agents often use heuristics

Perfect **information** and **computing** capabilities
Agents will learn the **"true model"**
Representative agent

No true model, agents can switch

Agents have a variety of behaviours and beliefs

Favourite-longshot bias

- Definition
  - Outcomes with high probability are underpriced 高概率结果的定价偏低
  - Outcomes with low probability are overpriced 概率低的结果定价过高
- **Main explanations**
  - "Insider" trading "内幕 "交易
    - Two types of agents. Random and insiders.
    - Insiders know the real probability.
  - Misperception of probabilities 对概率的误解
    - High values are underestimated 高值被低估
    - low values are overestimated 低值被高估
  - Risk-love
    - Some people are averse to risk

        - Some people seek risk

- **Prediction markets**
    - π_i Price of ticket to bet on i
    - p_i True Probability of i happening

  Pays 1 if i occurs, 0 if it doesn't, So you can win $1 - \pi_i$ or lose π_i

- **ABM 风险态度（Risk Attitude）**

  From prospect theory - our value/utility

$$v(x) = \begin{cases} x^\alpha & \text{if } x \geq 0 \\ -(-x)^\alpha & \text{if } x < 0 \end{cases}$$

$$w(p) = e^{-[-ln(p)]^\beta}$$

  From research on probability misperception - our expected probability

$$\boxed{u(\pi_i, p_i) = w(p_i)v(1 - \pi_i) + w(1 - p_i)v(-\pi_i)}$$

    - **Random**

      No "function", bet on a or b randomly (50/50 chance)
    - **Insiders**

      $\alpha = \beta = 1$
    - **Risk averse**

      $\alpha = 0.5$ $\beta = 1$
    - **Misperceiving agents**

      $\alpha = 1$ $\beta = 0.928$
    - **Risk lovers**

      $\alpha = 2$ $\beta = 1$

Risk averse

$$\pi^s(p) = \frac{p - \sqrt{p - p^2}}{2p - 1}$$

Risk lovers

$$\pi^s(p) = \frac{p^2}{1 - 2p + 2p^2}$$

Financial market

- Basic financial abm
    - **Chartists 宪章主义者 :** Follow the trend Short-term investors

$$s^e_{t+1} = (1-g)s_t + g s_{t-1}$$

Expected price **Price at time t** **Price at time t-1**

Short-term investors

$$\boxed{\Delta s^e_{t+1} = -g\Delta s_t}$$

- **Fundamentalists 原教旨主义者 :** Derive the "fundamental" price Long-term investors

$$s^e_{t+1} = (1-v)s_t + v\bar{s}_t$$

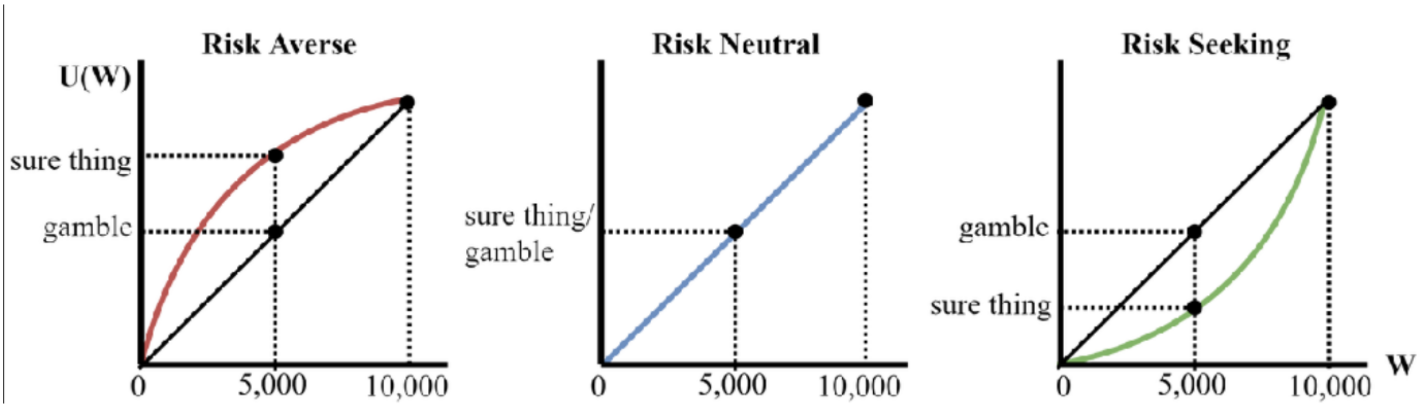
Expected price **Price at time t** **Fundamental Price at time t**

Long-term investors

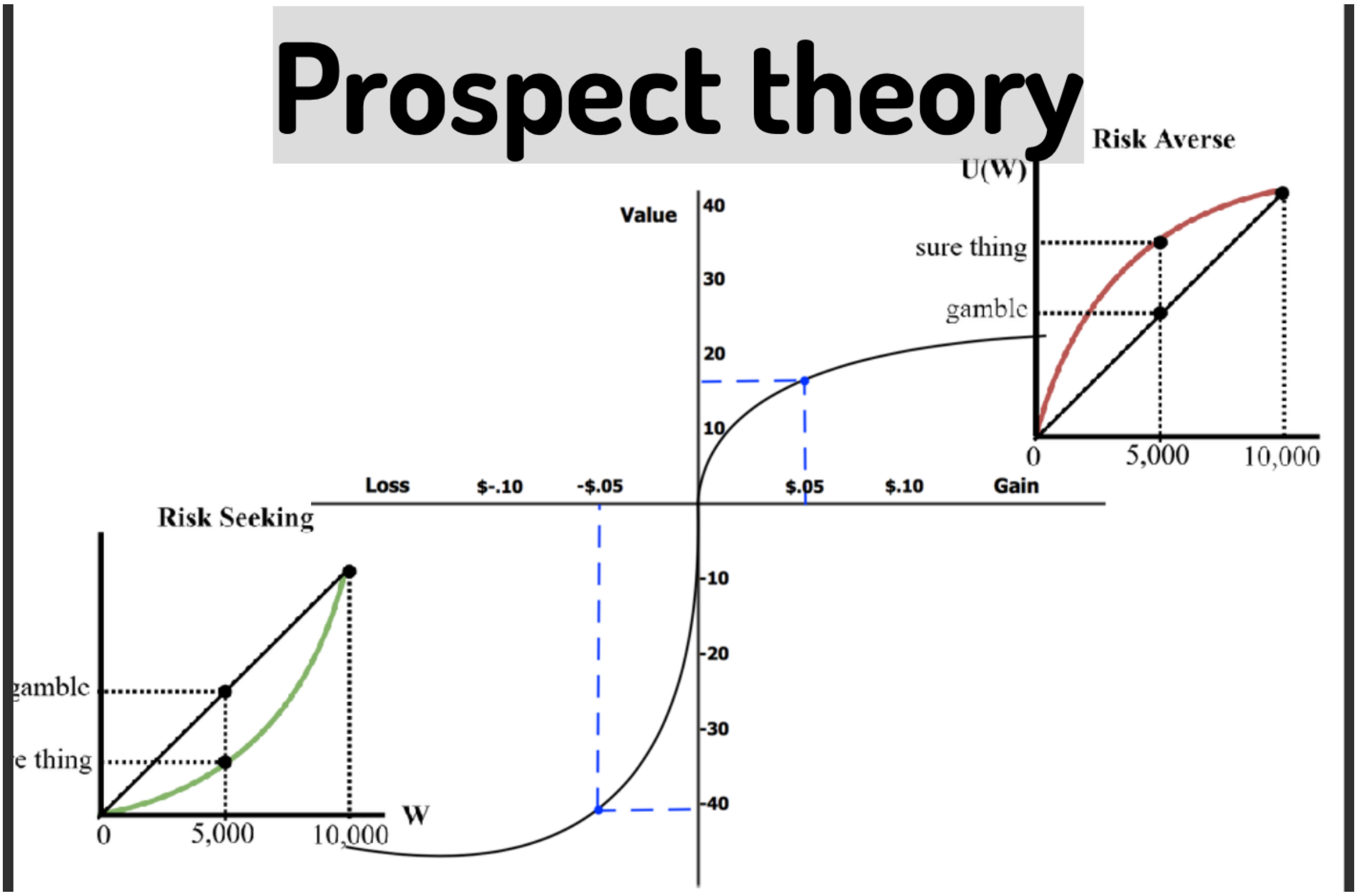
$$\boxed{\Delta s^e_{t+1} = v(\bar{s}_t - s_t)}$$

- Switching behaviour
  - Agents can compare different heuristics and choose
  - Agents can change their behaviour

Utility function



Prospect theory



$$v(x) = \begin{cases} x^{\alpha} & if\ x \geq 0 \\ -\lambda(-x)^{\beta} & if\ \ x < 0 \end{cases}$$

risk attitude（a,b） and loss aversion coefficient（λ）

Tversky and Kahneman

- **Availability**
  - We tend to overestimate what’s “available" in our memory 我们往往会高估记忆中的 "可用 "内容
- **Representativeness 代表性**
  - We fail at computing conditional probabilities 我们计算条件概率失败
- **Anchoring 固定**
  - We often choose a reference point 我们通常会选择一个参照点

The adaptive toolbox

- **Psychologic plausibility 心理合理性**
  - The aim is to build a model that accurately represents the behaviour of humans 目的是建立一个能准确代表人类行为的模型
- **Domain specific 特定领域**
  - Heuristics should be specific to the context, rather than general 启发式方法应针对具体情况，而不是泛泛而谈
- **Ecological rationality 生态理性**
  - The success of the heuristic is based on adaptation to the environment 启发式的成功取决于对环境的适应性

- **Heuristics**
  - Search rules
  - Stopping rules
  - Decision rules

Hyperbolic discounting 双曲贴现

- We are not good ad judging time 我们不善于判断时间
- We want everything now
- Instant gratification 及时行乐
- **Classical economics:**

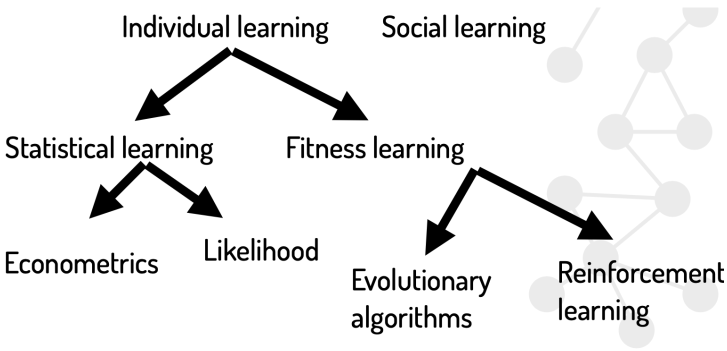
$$\frac{1}{1+k}^{t}$$

- **Reality (Behavioural economics)**

$$\frac{1}{1+kt}$$

- Learning
  - Agents have a **limited** or even a **wrong comprehension** of their environment代理对环境的理解有限甚至错误
  - They master **only a subset of all the actions** that can be conceived in order to face a given situation 它们只是面对特定情况时可以设想的所有行动的一个子集
  - They have an **imprecise understanding** of their own **goals** and **preferences**. 他们对自己的目标和偏好理解不准确。

  - **Objects of Learning**
    - Models of the world
    - Parameters within a given model
    - Actions
    - Realised outcomes
  - Types of Learning



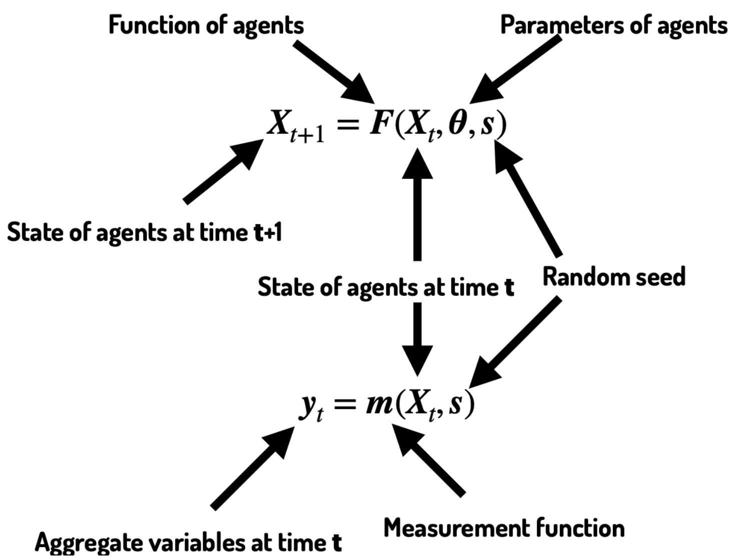
Interaction

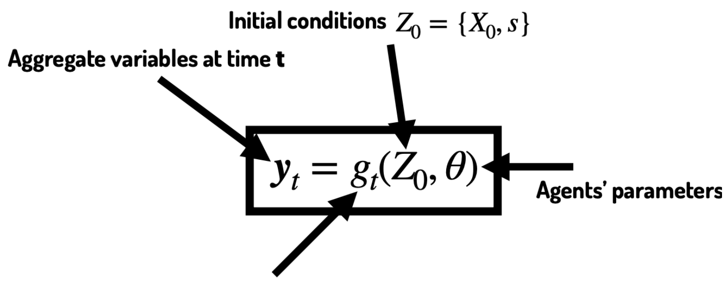
- Traditional economic theory overlooks interactions
- Mathematical modelling of interactions is complicated 相互作用的数学建模非常复杂
- ABMs are perfect for this

- **Local vs global**
  - **Local** if agents interact with "neighbourhood"
  - **Global**, agents take into account all the population
- **Exogenous vs endogenous 外源性与内源性**
  - **Exogenous** if don't change over time
  - **Endogenous** if agents can choose interactions
- **Deterministic vs stochastic 确定性与随机性**
  - **Deterministic** if interactions are between same agents
  - **Stochastic** if interaction involves probability
- **Direct vs indirect**
  - **Direct** if agents are affected directly by interaction 如果代理人直接受到互动的影响，则为直接影响
  - **Indirect** if interactions affect model only (but the model affects the agents somehow) 如果互动只影响模型，则为间接互动（但模型会对代理产生某种影响）

Equilibrium 平衡

- Equilibrium in abms can only be defined at the aggregate level 总体水平
- **Transition equation and Measurement equation:**





$$g_t(Z_0, \theta) = m(F(F \ldots (F(Z_0, \theta) \ldots))) = m(F^t(Z_0, \theta))$$

For each statistics , if statistical equilibrium is achieved in a given time window , then is stationary 对于每个统计量，如果在给定的时间窗口内实现了统计均衡，那么就是静止的

- **Ergodicity 各态历经**

yt is **ergodic** if the model reaches the **same** equilibria **irrespective of the initial conditions 如果模型无论初始条件如何都能达到相同的均衡点，则 yt 具有遍历性**

- **Sensitivity analysis 敏感性分析**
  - To understand which variables are useful
  - To understand the impact of each variable on the output
  - **Settings**
    - Factor screening
      - Identify the most influential factors
    - Local sensitivity analysis
      - Fine grained search around predetermined values 围绕预定值进行精细搜索
    - Global sensitivity analysis
      - Search around the entire range of values
  - **Strategies**
    - Factor prioritisation 因素优先排序
    - Factor fixing
      - Fixes factors that do not contribute too much to output
    - Factor mapping
      - Focus on critical areas (thresholds, phase transitions, etc.)
    - Metamodelling
      - To identify the relation between input and output

Validation

- **Define:**

  the **validity** of a model can be defined as the degree of **homomorphism** between a certain system (the model) and another system that it purportedly represents (the real-world system) 模型的有效性可以定义为某个系统（模型）和它声称代表的另一个系统（真实世界系统）之间的同构程度

- **Dimensions:**
  - Concept validation 概念验证
  - Empirical validation 经验验证
    - **Input validation**
      - Parameters and initial conditions 参数和初始条件
      - Structural assumptions 结构假设
    - **Output validation**
      - *Use as much data as you can!!!*

  Methodological basis of empirical validation 经验验证的方法论基础

    - Real-world data-generating process (**R**) vs model data-generating process (**M**)
    - Assessment
      - **Useful** if exhibit **at least some** of the observed historical patterns 如果至少表现出一些观察到的历史模式，则是有用的
      - Accurate if it exhibits only patterns observed historically 如果它只展示历史上观察到的模式，则是准确的
      - Complete if it exhibits all historical observed patterns 如果展示了所有历史观察到的模式，则为完整模式

$R \cap M = \varnothing$	Useless
$M \subset R$	Incomplete
$R \subset M$	Inaccurate/redundant
$M \Longleftrightarrow R$	Complete and accurate

- **Tractability vs accuracy 可操作性与准确性**
  - Simplification 简化
  - Stylised facts 风格化的事实
  - Trade-off between tractability and accuracy 可操作性与准确性之间的权衡
- **Instrumentalism vs realism 工具主义与现实主义**
  - Realism says theory is real 现实主义认为理论是真实的
  - Instrumentalism says theory is an instrument 工具论认为理论是一种工具
  - Are we interested in reproducing the world or in making predictions? 我们是对再现世界感兴趣，还是对进行预测感兴趣?
- **Problems of ABMs**

  ABMs are flexible and allow us to capture the complexity of a system. However, because of this, we need more data for validation. More data also means more complexity. This can cause issues such as overfitting and lack of interpretability.

  ABM 非常灵活，可以让我们捕捉到系统的复杂性。然而，正因为如此，我们需要更多的数据来进行验证。更多的数据也意味着更多的复杂性。这会导致过度拟合和缺乏可解释性等问题。

Calibration vs estimation 校准与估算

- **Calibration**: minimise the difference between R and M
- **Estimation**: find the true values of the variables

Compare the **model-produced data** with the **real-world-produced data**?

- **Preliminary tests 初步测试**
  - **Stationarity** is easy to test but things can be tricky if the data is not stationary
  - **Ergodicity** 遍历性 cannot be tested unless we observe many realisations of the same process
- **Method of simulated moments**
  - Minimise the distance between the summary statistics of the real-world system and those of our model 最小化真实世界系统的汇总统计与我们模型的汇总统计之间的距离
  - Method:

$$\hat{\theta} = argmin_{\theta}[\mu^*(\theta) - \mu_R]' W^{-1} [\mu^*(\theta) - \mu_R]$$

Weights

- **Bayesian estimation 贝叶斯估计**

  More appropriate when:

  - inference is needed 需要推理
  - there are prior and posterior distributions 有先验分布和后验分布

Abms cheat sheet

Agent design

1. **Nature** of agents
2. **List of variables describing their state**
3. List of **actions** the agents can perform
4. Structure of their interaction with other agents

Experiment

1. Definition of output variables of interest 相关产出变量的定义
2. Appropriate experimental design 适当的实验设计
3. Analysis of equilibria 平衡分析
4. Sensitivity analysis 敏感性分析

Validation:

1. Select the appropriate data 选择适当的数据
2. Input validation 输入验证
3. Output validation 输出验证

Estimation 评估

1. Select the appropriate data 选择适当的数据
2. Test for stationarity and ergodicity 静止性和遍历性检验
3. Methods of simulated distances 模拟距离的方法

Stylised facts 风格化的事实

是在金融时间序列中观察到的经验性规律

- Empirical Regularities 经验规律
  - patterns that have been observed so many times they are accepted as truth 多次观察到的模式被视为真理
- Returns
  - Price **cannot be compared** across different stocks 无法比较不同股票的价格
  - Price time series display trends 价格时间序列显示趋势
  - **Solution**: log-returns (difference of log of price)

***S*(*t*)**	Price at time t
***X*(*t*) = *ln*(*S*(*t*))**	Log price
Δ*t*	Time unit (minutes, hour, day, etc.)
***r*(*t*, Δ*t*) = *X*(*t* + Δ*t*)– *X*(*t*)**	Log return

$$r(t, \Delta t) = X(t + \Delta t) - X(t)$$

Stationary
Time invariant 时间不变量
Approximate relative returns well when returns are small 当收益较小时，能很好地近似相对收益

- **Autocorrelation 自相关**

  $C(\tau) = corr[r(t, \Delta t), r(t + \tau, \Delta t)]$

  $C(\tau) = 0, \forall \tau$

- Distribution of returns

$$\kappa[X] = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$$

$\kappa[X] = 0$	Normal distribution
$\kappa[X] > 0$	Long tails

- **Aggregational gaussianity 聚合高斯性**

$$\lim_{\Delta t \to \infty} \kappa = 0$$

- **Gain/loss Asymmetry 损益不对称**
  - Downwards movements are larger but fewer
  - Upwards movements are more frequent but smaller
  - This does not apply to forex 这不适用于外汇
- **Volume**
  - Volume is correlated with all measures of volatility 交易量与所有波动率指标都有关联
- **Calendar effects**
  - Price and volume display regularities based on day of week, week of month, month of year, etc. 根据星期、月份、月份等显示价格和成交量的规律。
- **Long memory**

$$\lim_{\tau \to \infty} \frac{\overset{\textbf{Correlation}}{C(\tau)}}{\underbrace{c_\tau}_{\textbf{Positive costant}}\,\underbrace{\tau^{-\alpha}}_{\alpha \in [0,1]}} = 1$$

$$H = 1 - \frac{\alpha}{2}$$

$$\frac{1}{2} < H < 1$$

  - Prices were found to have long memory, until they didn't. This led to a more robust reformulation of this statistic by Lo (1991). 人们发现价格具有长期记忆，直到它们不具有长期记忆。这促使 Lo（1991 年）对这一统计量进行了更稳健的重新表述。
  - Volume, volatility, bid/ask spread were found to exhibit long memory across several different markets and timescales by a number of studies (albeit with some differences) 许多研究发现，成交量、波动率、买入价/卖出价差在多个不同市场和时间尺度上表现出长记忆（尽管存在一些差异）。

Financial markets model

Lux and Marchesi 1999

- Considered one of the best early examples of ABM for financial markets
- Two types of agents, switching behaviour
- Many following abms were inspired by this one

Model:

- Fundamentalists 原教旨主义者
- Noise traders (chartists)
    - Optimistic
    - Pessimistic

Switching behaviour between optimistic and pessimistic traders

Switching between fundamentalists and chartists

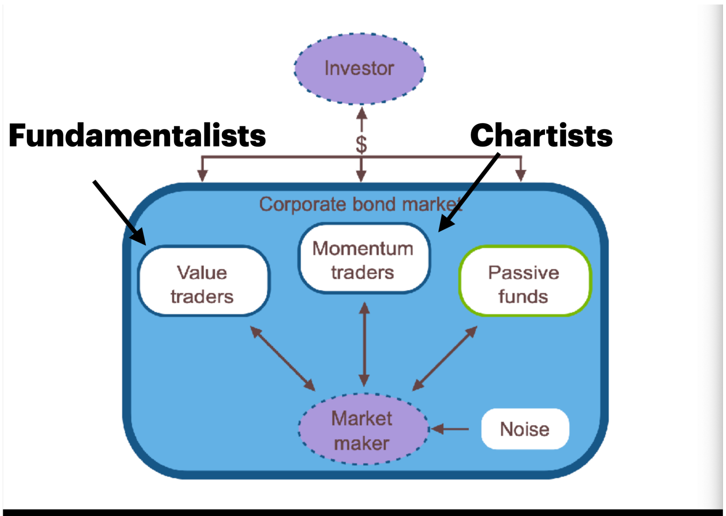
Probabilistic switching function based on profit differential

Robustness:

- For “many different parameter sets”, they find that the distribution of returns shows fat tails, and volatility exhibits long memory 对于 "许多不同的参数集"，他们发现回报率的分布呈现肥尾，波动率表现出长记忆性

Bond trading

- Model by Bank of England
- Model to explore the impact of shocks
- Findings to aid policy making



Braun-Munzinger, Liu, Turrell "An agent-based model of dynamics in corporate bond trading" - BoE working paper - 2016

They find that managing bond redemption by spreading payments over several days can significantly reduce shocks 他们发现，通过将付款分散到几天内来管理债券赎回，可以大大减少冲击

Credit risk - housing market

- Model housing market with heterogeneous agents 具有异质代理的住房市场模型
- They include lenders as well as households 它们包括贷款人和家庭

- They develop a policy to attenuate price cycles 他们制定了一项减缓价格周期的政策

Validation

They show that **aggregate variables** are correlated as expected in every scenario —> the model is consistent and reproduces empirical observations 结果表明，总体变量在每种情况下都如预期般相互关联 -> 模型是一致的，并再现了经验观察结果

Experiments

They perform a number of experiments to study the effect of various policies on 9 different key housing market indicators 他们进行了一系列实验，研究各种政策对 9 个不同的主要住房市场指标的影响

Results

Among other things, they find that an increase in the size of the buy-to-let markets leads to amplification of price cycles 除其他外，他们还发现购房出租市场规模的扩大会导致价格周期的放大

Public health

Contagion model (starting from standard sir) 传染模型（从标准先生开始）

Use of multiplex networks 使用多路复用网络

Develop non-pharmaceutical interventions 制定非药物干预措施