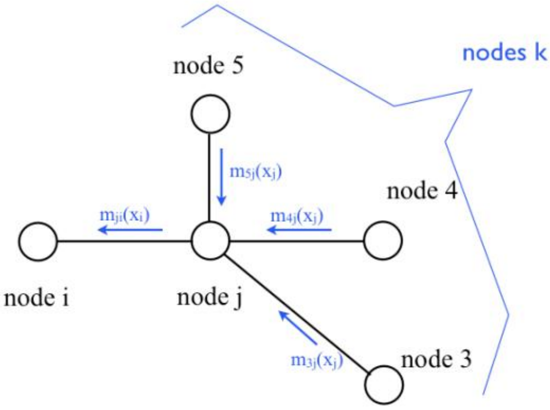
Belief Propagation Algorithm

效果：local computation => global inference

1. Acyclic

Essence: ordered 计算（global：hierarchical position => global inference）



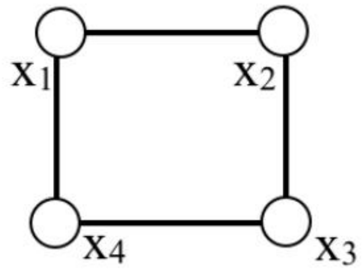


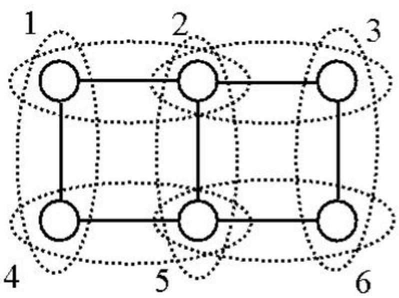
Figure 7.17: A loopy graph

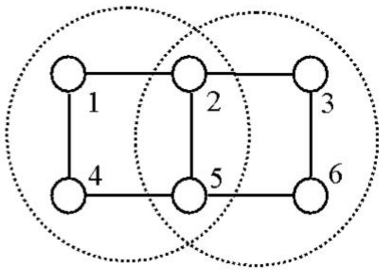
2. Cyclic

操作：引入 belief 和 message 的概念，通过 pairwise 的计算和迭代，得到 marginal distribution

Essence: 简化地认为局部（pairwise）的关系足以表示全局（Bethe approximation）

Key approximation:
$$b(\{x\}) = \frac{\prod_{(ij)} b_{ij}(x_i, x_j)}{\prod_i b_i(x_i)^{q_i - 1}}$$





Why works?

迭代更新的 converge 条件 <=> Bethe free energy 的最小值条件 （证明：normalization 的条件作为 Lagrange restriction terms）

Free Energy Principal

Origin: Physics => Neuroscience（Helmholtz）

Main idea: 两个 models (distributions)，最小化 KL divergence

=> neuroscience: 脑内模型（ensemble density q(θ; λ)) vs 基于 sensory input/观察值的θ的 conditional distribution （p(θ|y(α))）

θ: external causes 脑内世界模型的参数

λ: internal structures e.g.神经连接结构，神经激发情况

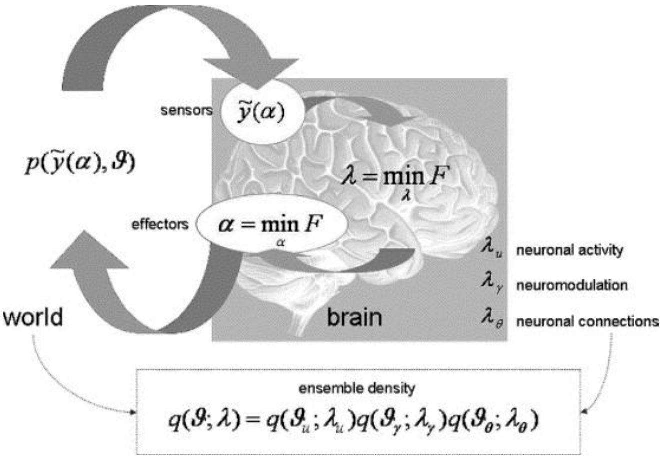
y: sensory input 感性 e.g. brain wave　　α: action

Free energy= KL divergence - model evidence = D{q(θ; λ)|| p(θ|y(α)}- lnp(y(α))

$$D\left(b\{x\}||p(\{x\}\right)=\sum_{\{x\}}b(\{x\})E(\{x\})+\sum_{\{x\}}b(\{x\})\ln b(\{x\})+\ln Z$$

$$G\left(b(\{x\})\right)=\sum_{\{x\}}b(\{x\})E(\{x\})+\sum_{\{x\}}b(\{x\})\ln b(\{x\})=U(b\{x\})-S(b\{x\})$$

Comparison: Minimize 是双向、交替进行的



$$F=-\int q(\vartheta)\ln\frac{p(\tilde{y},\vartheta)}{q(\vartheta)}\mathrm{d}\vartheta=-\langle\ln p(\tilde{y},\vartheta)\rangle_q+\langle\ln q(\vartheta)\rangle_q$$

t 时的脑内模型 q_t(θ; λ) & sensory input y_t(α_t-1) => 调整λ来 minimize F

$$F=-\ln p(\tilde{y})+\boxed{D(q(\vartheta;\lambda)\|p(\vartheta|\tilde{y}))}$$

得到 t+1 时刻的脑内模型 q_t+1(θ; λ)（本质是对 p_t(θ|y_t(α_t-1))的 approximation）
=> 根据模型 q_t+1(θ; λ)指导下一步行动，即α_t 来 minimize F（maximize p_t+1(y(α_t)|θ)）

$$F=\boxed{-\langle\ln p(\tilde{y}(\alpha)|\vartheta)\rangle_q}+D(q(\vartheta)\|p(\vartheta))$$

=> ......
不断根据行动结果调整脑内模型，根据模型进行“最优”行动

Ideas

Rough ideas:
BP 算法：已知最终目标是实现 Free energy 最小化；不一定收敛（与 topology 的 loopy 程度有关， entropy term 的 convex or not）
1. Initialization：topology；初始分布(equal or extreme) => 收敛？收敛速度？
2. 过程：是否有 patterns e.g.,
topology/network 边缘的 belief 和 message 更新快（是否也暗示了 global property）？
更新方向，extreme 的越 extreme，equal 越 equal？是否与 position 有关？
3. 过程（收敛前进行人为干涉）
1) 改变 message（e.g., 在调整认知过程中，即收敛前，受到新刺激，获得 extra 信息）

2) 新 structure：topology（e.g., 敲掉/增加特定节点；改变 overall 结构）；分布（e.g. extreme=>equal，比如突然没理由地随机改变认知模型，突变算法？）

4. 已知目标（某 marginal distribution），能设计一次收敛吗？（e.g., Initialization, topology，中途干涉步骤...）

<u>Interest</u>: 意识的运作, e.g., modular theory 模块之间如何 self-organize