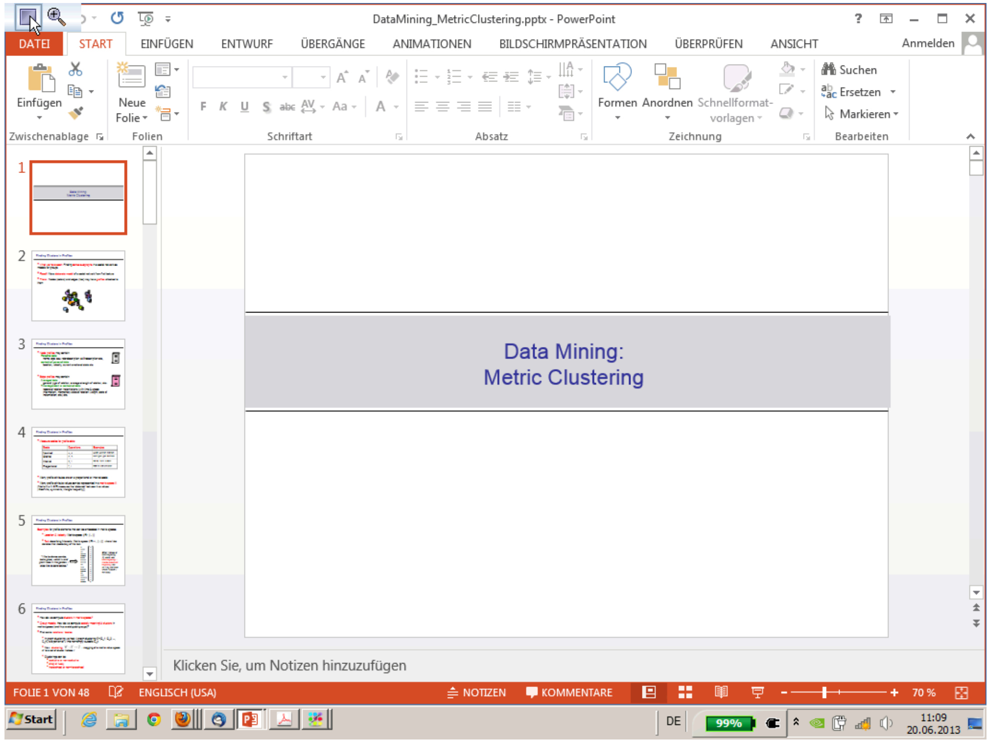
Script **generated by TTT**

Title: profile1 (20.06.2013)

Date: Thu Jun 20 11:09:16 CEST 2013

Duration: 92:32 min

Pages: 61



GMM-Basics

Maximum likelihood (one multivariate Gaussian)

$$p(\mathbf{x}|\theta) = \mathcal{N}(\mathbf{x}|\underbrace{\boldsymbol{\mu}, \boldsymbol{\Sigma}}_{\theta}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

- Likelihood $L(\mathbf{x},\theta) = p(\mathbf{x}|\theta)$
- Maximum likelihood $\theta_{\text{best}} = \operatorname{argmax}_{\theta} L(\mathbf{x},\theta) = \operatorname{argmax}_{\theta} \ln L(\mathbf{x},\theta)$
- Pattern matrix $\mathbf{X}$ of N iid measurements (D-dim. pattern vectors $\mathbf{x}$),

$$\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)^{\mathrm{T}}$$

$$L(\mathbf{X},\Theta) = \prod_{i=1}^{N} L(\mathbf{x}_i,\Theta) \qquad \ln L(\mathbf{X},\Theta) = \sum_{i=1}^{N} \ln L(\mathbf{x}_i,\Theta)$$

$$\ln L(\mathbf{X},\Theta) = \ln p(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \sum_{i=1}^{N} \ln N(\mathbf{x}_i|\boldsymbol{\mu},\boldsymbol{\Sigma})$$

GMM-Basics

Maximum likelihood (one multivariate Gaussian)

$$p(\mathbf{x}|\theta) = \mathcal{N}(\mathbf{x}|\underbrace{\boldsymbol{\mu}, \boldsymbol{\Sigma}}_{\theta}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

- Likelihood $L(\mathbf{x},\theta) = p(\mathbf{x}|\theta)$
- Maximum likelihood $\theta_{\text{best}} = \operatorname{argmax}_{\theta} L(\mathbf{x},\theta) = \operatorname{argmax}_{\theta} \ln L(\mathbf{x},\theta)$
- Pattern matrix $\mathbf{X}$ of N iid measurements (D-dim. pattern vectors $\mathbf{x}$),

$$\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)^{\mathrm{T}}$$

$$L(\mathbf{X},\Theta) = \prod_{i=1}^{N} L(\mathbf{x}_i,\Theta) \qquad \ln L(\mathbf{X},\Theta) = \sum_{i=1}^{N} \ln L(\mathbf{x}_i,\Theta)$$

$$\ln L(\mathbf{X},\Theta) = \ln p(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \sum_{i=1}^{N} \ln N(\mathbf{x}_i|\boldsymbol{\mu},\boldsymbol{\Sigma})$$

GMM-Basics

Maximum likelihood (one multivariate Gaussian)

$$p(\mathbf{x}|\theta) = \mathcal{N}(\mathbf{x}|\underbrace{\boldsymbol{\mu}, \boldsymbol{\Sigma}}_{\theta}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

- Likelihood $L(\mathbf{x},\theta) = p(\mathbf{x}|\theta)$

- Maximum likelihood $\theta_{\text{best}} = \operatorname{argmax}_{\theta} L(\mathbf{x},\theta) = \operatorname{argmax}_{\theta} \ln L(\mathbf{x},\theta)$

- Pattern matrix X of N iid measurements (D-dim. pattern vectors $\mathbf{x}$),

$$\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)^{\mathrm{T}}$$

$$L(\mathbf{X},\Theta) = \prod_{i=1}^{N} L(\mathbf{x}_i,\Theta) \qquad \ln L(\mathbf{X},\Theta) = \sum_{i=1}^{N} \ln L(\mathbf{x}_i,\Theta)$$

$$\ln L(\mathbf{X},\Theta) = \ln p(\mathbf{X}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma}) = \sum_{i=1}^{N} \ln N(\mathbf{x}_i\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma})$$

GMM-Basics

Maximum likelihood (one multivariate Gaussian)

$$p(\mathbf{x}|\theta) = \mathcal{N}(\mathbf{x}|\underbrace{\boldsymbol{\mu}, \boldsymbol{\Sigma}}_{\theta}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

- Likelihood $L(\mathbf{x},\theta) = p(\mathbf{x}|\theta)$

- Maximum likelihood $\theta_{\text{best}} = \operatorname{argmax}_{\theta} L(\mathbf{x},\theta) = \operatorname{argmax}_{\theta} \ln L(\mathbf{x},\theta)$

- Pattern matrix X of N iid measurements (D-dim. pattern vectors $\mathbf{x}$),

$$\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)^{\mathrm{T}}$$

$$L(\mathbf{X},\Theta) = \prod_{i=1}^{N} L(\mathbf{x}_i,\Theta) \qquad \ln L(\mathbf{X},\Theta) = \sum_{i=1}^{N} \ln L(\mathbf{x}_i,\Theta)$$

$$\ln L(\mathbf{X},\Theta) = \ln p(\mathbf{X}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma}) = \sum_{i=1}^{N} \ln N(\mathbf{x}_i\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma})$$

GMM-Basics

Maximum likelihood (one multivariate Gaussian)

$$\ln L(\mathbf{X},\Theta) = \ln L(\mathbf{X},\boldsymbol{\mu},\boldsymbol{\Sigma}) =$$

$$\ln p(\mathbf{X}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_n-\boldsymbol{\mu})$$

$$\theta_{\text{best}} = \operatorname{argmax}_{\theta} \ln L(\mathbf{x},\theta) \rightarrow \quad \boldsymbol{\mu}_{\text{best}}: \frac{\partial}{\partial\boldsymbol{\mu}} \ln L(\mathbf{X},\boldsymbol{\mu},\boldsymbol{\Sigma}) = 0 \qquad \boldsymbol{\Sigma}_{\text{best}}: \frac{\partial}{\partial\boldsymbol{\Sigma}} \ln L(\mathbf{X},\boldsymbol{\mu},\boldsymbol{\Sigma}) = 0$$

$$\boldsymbol{\mu}_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n$$

$$\boldsymbol{\Sigma}_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n-\boldsymbol{\mu}_{\mathrm{ML}})(\mathbf{x}_n-\boldsymbol{\mu}_{\mathrm{ML}})^{\mathrm{T}}$$

GMM-Basics

- GMM

$$p(\mathbf{x}|\theta) = p(\mathbf{x}) = \sum_{k=1}^{K}\pi_k\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k). \qquad 0 \leqslant \pi_k \leqslant 1 \qquad \sum_{k=1}^{K}\pi_k = 1$$

- 1 of K representation

K-dimensional binary random variable $\mathbf{z}$

$z_k \in \{0,1\}$ and $\sum_k z_k = 1$

$p(z_k = 1) = \pi_k$

$$p(\mathbf{z}) = \prod_{k=1}^{K}\pi_k^{z_k}$$

- conditional probability

$$p(\mathbf{x}|z_k = 1) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k) \qquad p(\mathbf{x}|\mathbf{z}) = \prod_{k=1}^{K}\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)^{z_k}$$

$$p(\mathbf{x}) = \sum_{\mathbf{z}}\underbrace{p(\mathbf{z})p(\mathbf{x}|\mathbf{z})}_{p(x,z)} = \sum_{k=1}^{K}\pi_k\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$$

GMM-Basics

- **GMM** $p(\mathbf{x}|\theta) =$

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \qquad 0 \leqslant \pi_k \leqslant 1 \qquad \sum_{k=1}^{K} \pi_k = 1$$

- **1 of K representation**

K-dimensional binary random variable $\mathbf{z}$

$z_k \in \{0, 1\}$ and $\sum_k z_k = 1$

$$p(z_k = 1) = \pi_k$$

$$p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$$

- **conditional probability**

$$p(\mathbf{x}|z_k = 1) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \qquad p(\mathbf{x}|\mathbf{z}) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_k}$$

$$p(\mathbf{x}) = \sum_{\mathbf{z}} \underbrace{p(\mathbf{z})p(\mathbf{x}|\mathbf{z})}_{p(x,z)} = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

GMM-Basics

- **GMM**

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \qquad 0 \leqslant \pi_k \leqslant 1 \qquad \sum_{k=1}^{K} \pi_k = 1$$

- **1 of k representation**

K-dimensional binary random variable $\mathbf{z}$

$z_k \in \{0, 1\}$ and $\sum_k z_k = 1$

$$p(z_k = 1) = \pi_k$$

$$p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$$

remark: If we have several observations $\mathbf{x}_1, \ldots, \mathbf{x}_N$, then, because we have represented the marginal distribution in the form $p(\mathbf{x}) = \sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{z})$, it follows that for every observed data point $\mathbf{x}_n$ there is a corresponding latent variable $\mathbf{z}_n$.

$$p(\mathbf{x}) = \sum_{\mathbf{z}} \underbrace{p(\mathbf{z})p(\mathbf{x}|\mathbf{z})}_{p(x,z)} = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

GMM-Basics

Maximum likelihood (GMM)

$$\ln L(\mathbf{X}, \Theta) = \ln L(\mathbf{X}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\}$$

Vector of K D-dim. means $\boldsymbol{\mu}_k$

Vector of K DxD covariances $\boldsymbol{\Sigma}_k$

- maximizing w.r.t $\boldsymbol{\pi}$, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ →

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n \qquad \boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})(\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^{\mathrm{T}}$$

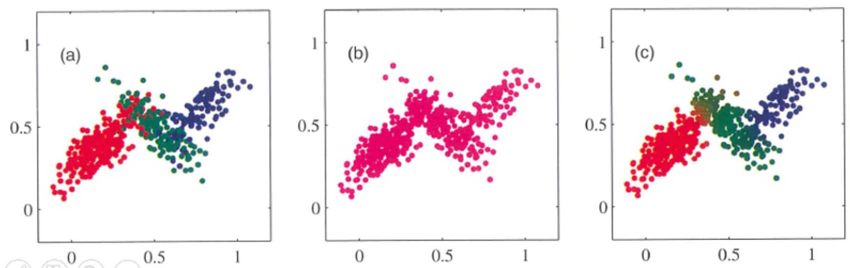
$$\left(N_k = \sum_{n=1}^{N} \gamma(z_{nk}). \right) \qquad \pi_k = \frac{N_k}{N}$$

GMM-Basics

- **Responsibilities**

$$\gamma(z_k) \equiv p(z_k = 1|\mathbf{x}) = \frac{p(z_k = 1)p(\mathbf{x}|z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)p(\mathbf{x}|z_j = 1)} = \frac{\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}.$$

- **Example**

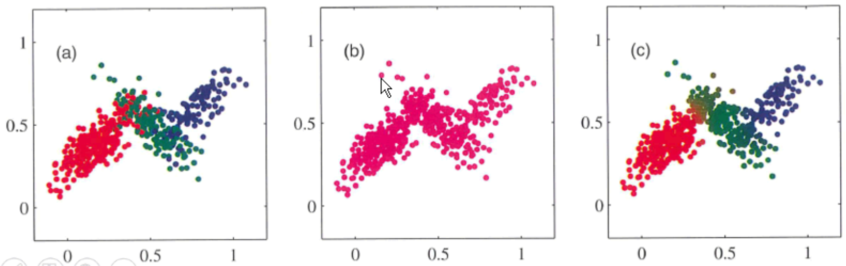


GMM-Basics

- Responsibilities

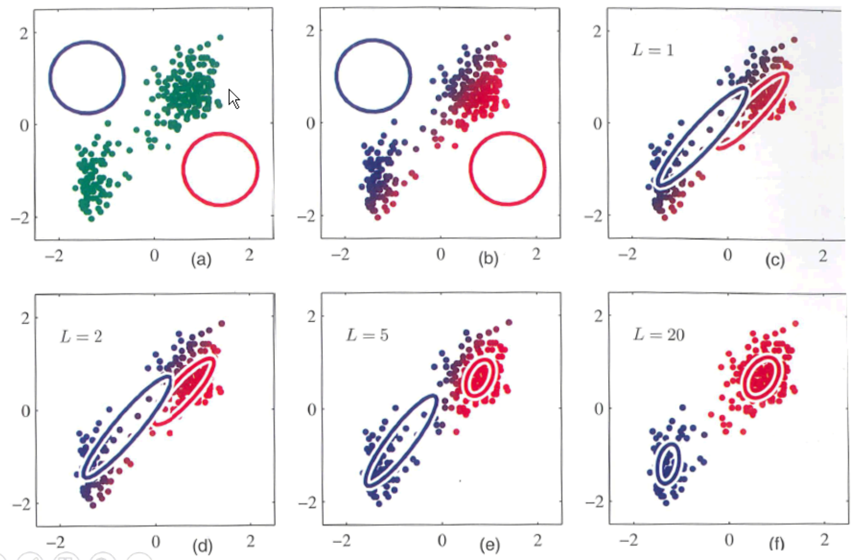
$$\gamma(z_k) \equiv p(z_k = 1|\mathbf{x}) = \frac{p(z_k = 1)p(\mathbf{x}|z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)p(\mathbf{x}|z_j = 1)} = \frac{\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}.$$

- Example



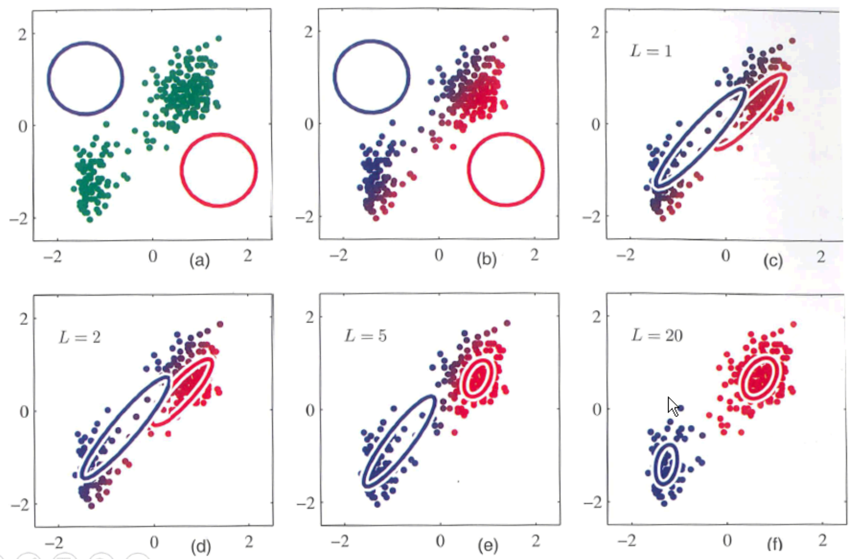
GMM-Basics

Maximum likelihood (GMM)



GMM-Basics

Maximum likelihood (GMM)



EM-algorithm: General View

- Having latent variables $\mathbf{Z}$, ML becomes

$$\ln p(\mathbf{X}|\boldsymbol{\theta}) = \ln \left\{ \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) \right\}$$

- Summation inside $\ln$ → Problems !

- If we knew the complete dataset $\{\mathbf{X}, \mathbf{Z}\}$ (and thus the distribution $p(\mathbf{X}, \mathbf{Z}|\theta)$), we could use ML to solve for θ with $p(\mathbf{X}, \mathbf{Z}|\theta)$ directly (which is easy, as we will see, because $p(\mathbf{X}, \mathbf{Z}|\theta)$ is of exponential family (the functional form is known!!)

- We only know $p(\mathbf{Z}|\mathbf{X},\theta)$ (→ responsibilities, as we will see) → compute expectation of (unknown) quantity $p(\mathbf{X}, \mathbf{Z}|\theta)$ or even better of the quantity $\ln p(\mathbf{X}, \mathbf{Z}|\theta)$

EM-algorithm: General View

- Having latent variables $\mathbf{Z}$, ML becomes

$$\ln p(\mathbf{X}|\boldsymbol{\theta}) = \ln\left\{\sum_{\mathbf{Z}} p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})\right\}$$

- Summation inside $\ln$ → Problems !

- If we knew the complete dataset $\{\mathbf{X},\mathbf{Z}\}$ (and thus the distribution $p(\mathbf{X},\mathbf{Z}|\theta)$), we could use ML to solve for θ with $p(\mathbf{X},\mathbf{Z}|\theta)$ directly (which is easy, as we will see, because $p(\mathbf{X},\mathbf{Z}|\theta)$ is of exponential family (the functional form is known!!)

- We only know $p(\mathbf{Z}|\mathbf{X},\theta)$ (→ responsibilities, as we will see) → compute expectation of (unknown) quantity $p(\mathbf{X},\mathbf{Z}|\theta)$ or even better of the quantity $\ln p(\mathbf{X},\mathbf{Z}|\theta)$

EM-algorithm: General View

- Having latent variables $\mathbf{Z}$, ML becomes

$$\ln p(\mathbf{X}|\boldsymbol{\theta}) = \ln\left\{\sum_{\mathbf{Z}} p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})\right\}$$

- Summation inside $\ln$ → Problems !

- If we knew the complete dataset $\{\mathbf{X},\mathbf{Z}\}$ (and thus the distribution $p(\mathbf{X},\mathbf{Z}|\theta)$), we could use ML to solve for θ with $p(\mathbf{X},\mathbf{Z}|\theta)$ directly (which is easy, as we will see, because $p(\mathbf{X},\mathbf{Z}|\theta)$ is of exponential family (the functional form is known!!)

- We only know $p(\mathbf{Z}|\mathbf{X},\theta)$ (→ responsibilities, as we will see) → compute expectation of (unknown) quantity $p(\mathbf{X},\mathbf{Z}|\theta)$ or even better of the quantity $\ln p(\mathbf{X},\mathbf{Z}|\theta)$

EM-algorithm: General View

- Having latent variables $\mathbf{Z}$, ML becomes

$$\ln p(\mathbf{X}|\boldsymbol{\theta}) = \ln\left\{\sum_{\mathbf{Z}} p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})\right\}$$

- Summation inside $\ln$ → Problems !

- If we knew the complete dataset $\{\mathbf{X},\mathbf{Z}\}$ (and thus the distribution $p(\mathbf{X},\mathbf{Z}|\theta)$), we could use ML to solve for θ with $p(\mathbf{X},\mathbf{Z}|\theta)$ directly (which is easy, as we will see, because $p(\mathbf{X},\mathbf{Z}|\theta)$ is of exponential family (the functional form is known!!)

- We only know $p(\mathbf{Z}|\mathbf{X},\theta)$ (→ responsibilities, as we will see) → compute expectation of (unknown) quantity $p(\mathbf{X},\mathbf{Z}|\theta)$ or even better of the quantity $\ln p(\mathbf{X},\mathbf{Z}|\theta)$

EM-algorithm: General View

- Having latent variables $\mathbf{Z}$, ML becomes

$$\ln p(\mathbf{X}|\boldsymbol{\theta}) = \ln\left\{\sum_{\mathbf{Z}} p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})\right\}$$

- Summation inside $\ln$ → Problems !

- If we knew the complete dataset $\{\mathbf{X},\mathbf{Z}\}$ (and thus the distribution $p(\mathbf{X},\mathbf{Z}|\theta)$), we could use ML to solve for θ with $p(\mathbf{X},\mathbf{Z}|\theta)$ directly (which is easy, as we will see, because $p(\mathbf{X},\mathbf{Z}|\theta)$ is of exponential family (the functional form is known!!)

- We only know $p(\mathbf{Z}|\mathbf{X},\theta)$ (→ responsibilities, as we will see) → compute expectation of (unknown) quantity $p(\mathbf{X},\mathbf{Z}|\theta)$ or even better of the quantity $\ln p(\mathbf{X},\mathbf{Z}|\theta)$

EM-algorithm: General View

- Having latent variables $\mathbf{Z}$, ML becomes

$$\ln p(\mathbf{X}|\boldsymbol{\theta}) = \ln \left\{ \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) \right\}$$

- Summation inside $\ln$ → Problems !

- If we knew the complete dataset $\{\mathbf{X}, \mathbf{Z}\}$ (and thus the distribution $p(\mathbf{X}, \mathbf{Z}|\theta)$), we could use ML to solve for θ with $p(\mathbf{X}, \mathbf{Z}|\theta)$ directly (which is easy, as we will see, because $p(\mathbf{X}, \mathbf{Z}|\theta)$ is of exponential family (the functional form is known!!)

- We only know $p(\mathbf{Z}|\mathbf{X},\theta)$ (→ responsibilities, as we will see) → compute expectation of (unknown) quantity $p(\mathbf{X}, \mathbf{Z}|\theta)$ or even better of the quantity $\ln p(\mathbf{X}, \mathbf{Z}|\theta)$

EM-algorithm: General View

- alternating EM:

E-Step: compute $$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}).$$

M-Step: compute $$\boldsymbol{\theta}^{\text{new}} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}).$$

EM-algorithm: General View

- alternating EM:

E-Step: compute $$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}).$$

M-Step: compute $$\boldsymbol{\theta}^{\text{new}} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}).$$

EM-algorithm: General View

- alternating EM:

E-Step: compute $$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}).$$

M-Step: compute $$\boldsymbol{\theta}^{\text{new}} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}).$$

EM-algorithm: General View

- alternating EM:

E-Step: compute
$$\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}).$$

M-Step: compute
$$\boldsymbol{\theta}^{\text{new}} = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}).$$

EM-algorithm: General View

- applied to GMM:

$$p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k} \qquad p(\mathbf{x}|\mathbf{z}) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_k} \quad \Rightarrow$$

$$p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) = \prod_{n=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{nk}} \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_{nk}}$$

$$\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) = \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \left\{ \ln \pi_k + \ln \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\}.$$

Bayes $\Rightarrow$
$$p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) \propto \prod_{n=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right]^{z_{nk}}$$

$$\frac{p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})}{p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})} = \frac{p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})}{\sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})}$$

EM-algorithm: General View

$$p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) \propto \prod_{n=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right]^{z_{nk}} \quad \Rightarrow$$

$$\begin{aligned}
\mathbb{E}[z_{nk}] &= \frac{\sum_{z_{nk} \in \{0,1\}} z_{nk} \left[\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right]^{z_{nk}}}{\sum_{z_{nj}} \left[\pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)\right]^{z_{nj}}} \\
&= \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} = \gamma(z_{nk})
\end{aligned}$$

EM: Relation to K-Means

- If we use k Gaussians with $\Sigma = \epsilon I$: *(that is why K-Means favors spherical clusters)*

$$p(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{(2\pi\epsilon)^{1/2}} \exp\left\{-\frac{1}{2\epsilon}\|\mathbf{x} - \boldsymbol{\mu}_k\|^2\right\}$$

- we get for the responsibilities:

$$\gamma(z_{nk}) = \frac{\pi_k \exp\left\{-\|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2/2\epsilon\right\}}{\sum_j \pi_j \exp\left\{-\|\mathbf{x}_n - \boldsymbol{\mu}_j\|^2/2\epsilon\right\}}$$

- Letting $\epsilon \rightarrow 0$ and Taylor-Expansion:

$$\mathbb{E}_{\mathbf{Z}}[\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})] \rightarrow -\frac{1}{2} \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2 + \text{const}$$

→ same as on slide 18

EM: Relation to K-Means

that is why K-Means favors spherical clusters

- If we use k Gaussians with $\Sigma = \epsilon I$:

$$p(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{(2\pi\epsilon)^{1/2}} \exp\left\{-\frac{1}{2\epsilon}\|\mathbf{x}-\boldsymbol{\mu}_k\|^2\right\}$$

- we get for the responsibilities:

$$\gamma(z_{nk}) = \frac{\pi_k \exp\{-\|\mathbf{x}_n-\boldsymbol{\mu}_k\|^2/2\epsilon\}}{\sum_j \pi_j \exp\{-\|\mathbf{x}_n-\boldsymbol{\mu}_j\|^2/2\epsilon\}}$$

- Letting ε → 0 and Taylor-Expansion:

$$\mathbb{E}_{\mathbf{Z}}[\ln p(\mathbf{X},\mathbf{Z}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\pi})] \to -\frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\|\mathbf{x}_n-\boldsymbol{\mu}_k\|^2 + \text{const}$$

→ same as on slide 18

Real World Networks: Properties and Models

Lecture will mostly follow [1], thus corresponding citations are often omitted to increase readability

Complex Network Properties

Now: investigate a series of properties / classification axes of complex real world networks (mostly compared to random NW)

Mean Average Path Length

- "Small World Effect": $l(n)$ "small" → $l(n) \in O(\log(n))$
- undirected graph:

$$\ell = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij}$$

formula also counts 0 distances from i to i: ½ n(n+1) = ½ n(n-1) + n

- Expression allowing for disconnected components (where $d_{ij}=\infty$ can occur): harmonic mean:

$$\ell^{-1} = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij}^{-1}$$

Mean Average Path Length

- "Small World Effect": $l(n)$ "small" → $l(n) \in O(\log(n))$
- undirected graph:

$$\ell = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij}$$

formula also counts 0 distances from i to i: ½ n(n+1) = ½ n(n-1) + n

- Expression allowing for disconnected components (where $d_{ij}=\infty$ can occur): harmonic mean:

$$\ell^{-1} = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij}^{-1}$$

Mean Average Path Length

- "Small World Effect": $l(n)$ "small" → $l(n) \in O(\log(n))$
- undirected graph:

$$\ell = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij}$$

formula also counts 0 distances from i to i: ½ n(n+1) = ½ n(n-1) + n

- Expression allowing for disconnected components (where $d_{ij}=\infty$ can occur): harmonic mean:

$$\ell^{-1} = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij}^{-1}$$

Transitivity / Clustering Coefficient

- Clustering coefficient (whole graph):

$$C = C^{(1)} = \frac{3 \times \text{ number of triangles in the network}}{\text{number of connected triples of vertices}} = \frac{6 \times \text{ number of triangles in the network}}{\text{number of paths of length two}}$$

p(FOAF)

- Clustering coefficient (Watts-Strogatz-version, for node i):

$$C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i} = \frac{|\{e_{\{kj\}} \mid v_k, v_j \in N_i\}|}{\frac{k_i(k_i-1)}{2}}$$

(see Introduction , k_i = degree of node i)

Clustering coefficient (Watts-Strogatz-version, for whole graph):

$$C = C^{(2)} = \frac{1}{n} \sum_i C_i$$

mean of ratio instead of ratio of means

Transitivity / Clustering Coefficient

- Clustering coefficient (whole graph):

$$C = C^{(1)} = \frac{3 \times \text{ number of triangles in the network}}{\text{number of connected triples of vertices}} = \frac{6 \times \text{ number of triangles in the network}}{\text{number of paths of length two}}$$

p(FOAF)

- Clustering coefficient (Watts-Strogatz-version, for node i):

$$C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i} = \frac{|\{e_{\{kj\}} \mid v_k, v_j \in N_i\}|}{\frac{k_i(k_i-1)}{2}}$$

(see Introduction , k_i = degree of node i)

Clustering coefficient (Watts-Strogatz-version, for whole graph):

$$C = C^{(2)} = \frac{1}{n} \sum_i C_i$$

mean of ratio instead of ratio of means

Transitivity / Clustering Coefficient

- Clustering coefficient (whole graph):

$$C = C^{(1)} = \frac{3\times \text{ number of triangles in the network}}{\text{number of connected triples of vertices}} = \frac{6\times \text{ number of triangles in the network}}{\text{number of paths of length two}}$$

p(FOAF)

- Clustering coefficient (Watts-Strogatz-version, for node i):

$$C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i} = \frac{|\{e_{\{kj\}} \mid v_k, v_j \in N_i\}|}{\frac{k_i(k_i-1)}{2}}$$

(see Introduction , k_i = degree of node i)

Clustering coefficient (Watts-Strogatz-version, for whole graph):

$$C = C^{(2)} = \frac{1}{n}\sum_i C_i$$

mean of ratio instead of ratio of means

Transitivity / Clustering Coefficient

- Clustering coefficient (whole graph):

$$C = C^{(1)} = \frac{3\times \text{ number of triangles in the network}}{\text{number of connected triples of vertices}} = \frac{6\times \text{ number of triangles in the network}}{\text{number of paths of length two}}$$

p(FOAF)

- Clustering coefficient (Watts-Strogatz-version, for node i):

$$C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i} = \frac{|\{e_{\{kj\}} \mid v_k, v_j \in N_i\}|}{\frac{k_i(k_i-1)}{2}}$$

(see Introduction , k_i = degree of node i)

Clustering coefficient (Watts-Strogatz-version, for whole graph):

$$C = C^{(2)} = \frac{1}{n}\sum_i C_i$$

mean of ratio instead of ratio of means

Transitivity / Clustering Coefficient

- Clustering coefficient (whole graph):

$$C = C^{(1)} = \frac{3\times \text{ number of triangles in the network}}{\text{number of connected triples of vertices}} = \frac{6\times \text{ number of triangles in the network}}{\text{number of paths of length two}}$$

p(FOAF)

- Clustering coefficient (Watts-Strogatz-version, for node i):

$$C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i} = \frac{|\{e_{\{kj\}} \mid v_k, v_j \in N_i\}|}{\frac{k_i(k_i-1)}{2}}$$

(see Introduction , k_i = degree of node i)

Clustering coefficient (Watts-Strogatz-version, for whole graph):

$$C = C^{(2)} = \frac{1}{n}\sum_i C_i$$

mean of ratio instead of ratio of means

Transitivity / Clustering Coefficient

- Clustering coefficient (whole graph):

$$C = C^{(1)} = \frac{3\times \text{ number of triangles in the network}}{\text{number of connected triples of vertices}} = \frac{6\times \text{ number of triangles in the network}}{\text{number of paths of length two}}$$

p(FOAF)

- Clustering coefficient (Watts-Strogatz-version, for node i):

$$C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i} = \frac{|\{e_{\{kj\}} \mid v_k, v_j \in N_i\}|}{\frac{k_i(k_i-1)}{2}}$$

(see Introduction , k_i = degree of node i)

Clustering coefficient (Watts-Strogatz-version, for whole graph):

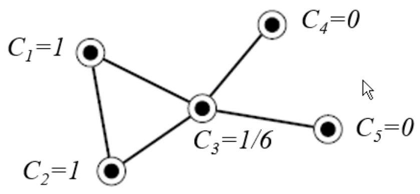
$$C = C^{(2)} = \frac{1}{n}\sum_i C_i$$

mean of ratio instead of ratio of means

Transitivity / Clustering Coefficient

Example:

$$C^{(1)} = \frac{3\times \text{ number of triangles in the network}}{\text{number of connected triples of vertices}} = \frac{3\times 1}{8} = 0.375$$



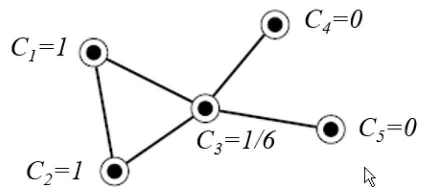
$$C^{(2)} = \frac{1}{n}\sum_i C_i \quad \text{with} \quad C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i}$$

$C^{(2)} = 1/5\ (1 + 1 + 1/6 + 0 + 0) = 13/30 = 0.433333$

Transitivity / Clustering Coefficient

Example:

$$C^{(1)} = \frac{3\times \text{ number of triangles in the network}}{\text{number of connected triples of vertices}} = \frac{3\times 1}{8} = 0.375$$



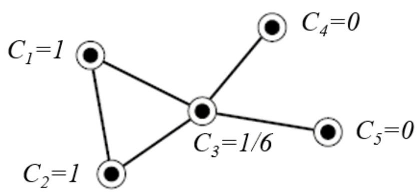
$$C^{(2)} = \frac{1}{n}\sum_i C_i \quad \text{with} \quad C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i}$$

$C^{(2)} = 1/5\ (1 + 1 + 1/6 + 0 + 0) = 13/30 = 0.433333$

Transitivity / Clustering Coefficient

Example:

$$C^{(1)} = \frac{3\times \text{ number of triangles in the network}}{\text{number of connected triples of vertices}} = \frac{3\times 1}{8} = 0.375$$



$$C^{(2)} = \frac{1}{n}\sum_i C_i \quad \text{with} \quad C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i}$$

$C^{(2)} = 1/5\ (1 + 1 + 1/6 + 0 + 0) = 13/30 = 0.433333$

	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	–0.029	45
	sexual contacts	undirected	2 810				3.2				265, 266
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	–0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	–0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	–0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	–0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	–0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	–0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	–0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	–0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	–0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	–0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	–0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	–0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	–0.226	416, 421

BLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices
ber of edges m; mean degree z; mean vertex–vertex distance ℓ; exponent α of degree distribution if the distribution follows a power law (or "–" if not; in/out
onents are given for directed graphs); clustering coefficient $C^{(1)}$ from Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r, Se
last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

[1]

	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	–0.029	45
	sexual contacts	undirected	2 810				3.2				265, 266
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	–0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	–0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	–0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	–0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	–0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	–0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	–0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	–0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	–0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	–0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	–0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	–0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	–0.226	416, 421

BLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices
ıber of edges m; mean degree z; mean vertex–vertex distance ℓ; exponent α of degree distribution if the distribution follows a power law (or "–" if not; in/out
onents are given for directed graphs); clustering coefficient $C^{(1)}$ from Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r, Sec
last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

[1]

	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	–0.029	45
	sexual contacts	undirected	2 810				3.2				265, 266
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	–0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	–0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	–0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	–0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	–0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	–0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	–0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	–0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	–0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	–0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	–0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	–0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	–0.226	416, 421

BLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices
ıber of edges m; mean degree z; mean vertex–vertex distance ℓ; exponent α of degree distribution if the distribution follows a power law (or "–" if not; in/out
onents are given for directed graphs); clustering coefficient $C^{(1)}$ from Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r, Sec
last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

[1]

	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	–0.029	45
	sexual contacts	undirected	2 810				3.2				265, 266
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	–0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	–0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	–0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	–0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	–0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	–0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	–0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	–0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	–0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	–0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	–0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	–0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	–0.226	416, 421

BLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices
ıber of edges m; mean degree z; mean vertex–vertex distance ℓ; exponent α of degree distribution if the distribution follows a power law (or "–" if not; in/out
onents are given for directed graphs); clustering coefficient $C^{(1)}$ from Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r, Sec
last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

[1]

	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	–0.029	45
	sexual contacts	undirected	2 810				3.2				265, 266
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	–0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	–0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	–0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	–0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	–0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	–0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	–0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	–0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	–0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	–0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	–0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	–0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	–0.226	416, 421

BLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices
ıber of edges m; mean degree z; mean vertex–vertex distance ℓ; exponent α of degree distribution if the distribution follows a power law (or "–" if not; in/out
onents are given for directed graphs); clustering coefficient $C^{(1)}$ from Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r, Sec
last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

[1]

	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	−0.029	45
	sexual contacts	undirected	2 810				3.2				265, 266
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	−0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	−0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	−0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	−0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	−0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	−0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	−0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	−0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	−0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	−0.226	416, 421

BLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices ber of edges m; mean degree z; mean vertex–vertex distance ℓ; exponent α of degree distribution if the distribution follows a power law (or "–" if not; in/out onents are given for directed graphs); clustering coefficient $C^{(1)}$ from Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r, Sec last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

[1]

	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	−0.029	45
	sexual contacts	undirected	2 810				3.2				265, 266
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	−0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	−0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	−0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	−0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	−0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	−0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	−0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	−0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	−0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	−0.226	416, 421

BLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices ber of edges m; mean degree z; mean vertex–vertex distance ℓ; exponent α of degree distribution if the distribution follows a power law (or "–" if not; in/out onents are given for directed graphs); clustering coefficient $C^{(1)}$ from Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r, Sec last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

[1]

	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	−0.029	45
	sexual contacts	undirected	2 810				3.2				265, 266
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	−0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	−0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	−0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	−0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	−0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	−0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	−0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	−0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	−0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	−0.226	416, 421

BLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices ber of edges m; mean degree z; mean vertex–vertex distance ℓ; exponent α of degree distribution if the distribution follows a power law (or "–" if not; in/out onents are given for directed graphs); clustering coefficient $C^{(1)}$ from Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r, Sec last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

[1]

Degree Distribution

- Notation:

  p(k) = p_k = fraction of nodes having degree k

- Cumulative distribution:

$$P_k = \sum_{k'=k}^{\infty} p_{k'}$$

- power law:

$$p_k \sim k^{-\alpha}$$

$$\rightarrow \quad P_k \sim \sum_{k'=k}^{\infty} k'^{-\alpha} \sim k^{-(\alpha-1)}$$

- exponential:

$$p_k \sim \mathrm{e}^{-k/\kappa}$$

$$\rightarrow \quad P_k = \sum_{k'=k}^{\infty} p_k \sim \sum_{k'=k}^{\infty} \mathrm{e}^{-k'/\kappa} \sim \mathrm{e}^{-k/\kappa}$$

Degree Distribution

- Notation:

  p(k) = p_k = fraction of nodes having degree k

- Cumulative distribution:

$$P_k = \sum_{k'=k}^{\infty} p_{k'}$$

- power law:

$$p_k \sim k^{-\alpha}$$

$$\rightarrow \quad P_k \sim \sum_{k'=k}^{\infty} k'^{-\alpha} \sim k^{-(\alpha-1)}$$

- exponential:

$$p_k \sim \mathrm{e}^{-k/\kappa}$$

$$\rightarrow \quad P_k = \sum_{k'=k}^{\infty} p_k \sim \sum_{k'=k}^{\infty} \mathrm{e}^{-k'/\kappa} \sim \mathrm{e}^{-k/\kappa}$$

Degree Distribution

- Notation:

  p(k) = p_k = fraction of nodes having degree k

- Cumulative distribution:

$$P_k = \sum_{k'=k}^{\infty} p_{k'}$$

- power law:

$$p_k \sim k^{-\alpha}$$

$$\rightarrow \quad P_k \sim \sum_{k'=k}^{\infty} k'^{-\alpha} \sim k^{-(\alpha-1)}$$

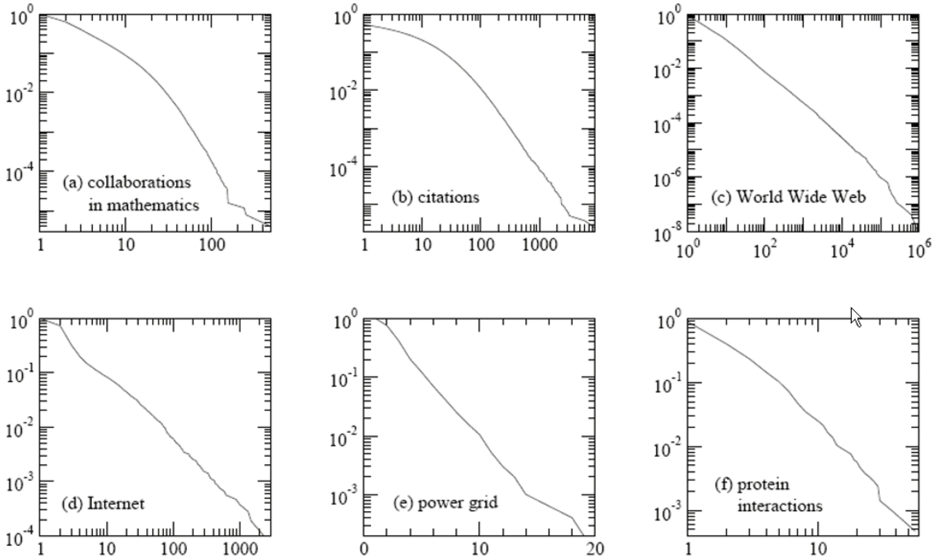
- exponential:

$$p_k \sim \mathrm{e}^{-k/\kappa}$$

$$\rightarrow \quad P_k = \sum_{k'=k}^{\infty} p_k \sim \sum_{k'=k}^{\infty} \mathrm{e}^{-k'/\kappa} \sim \mathrm{e}^{-k/\kappa}$$

Degree Distribution

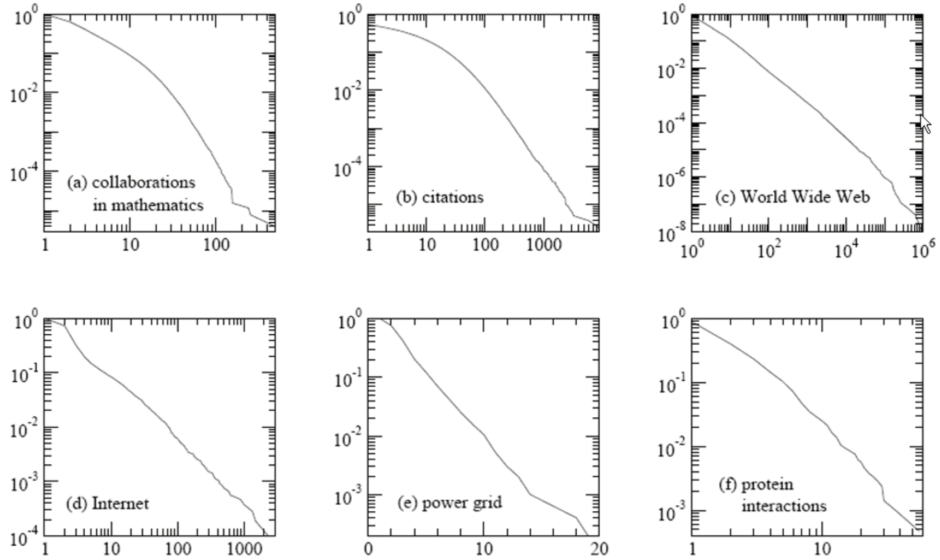
Cumulative distributions P_k of example real world NW



[1]

Degree Distribution

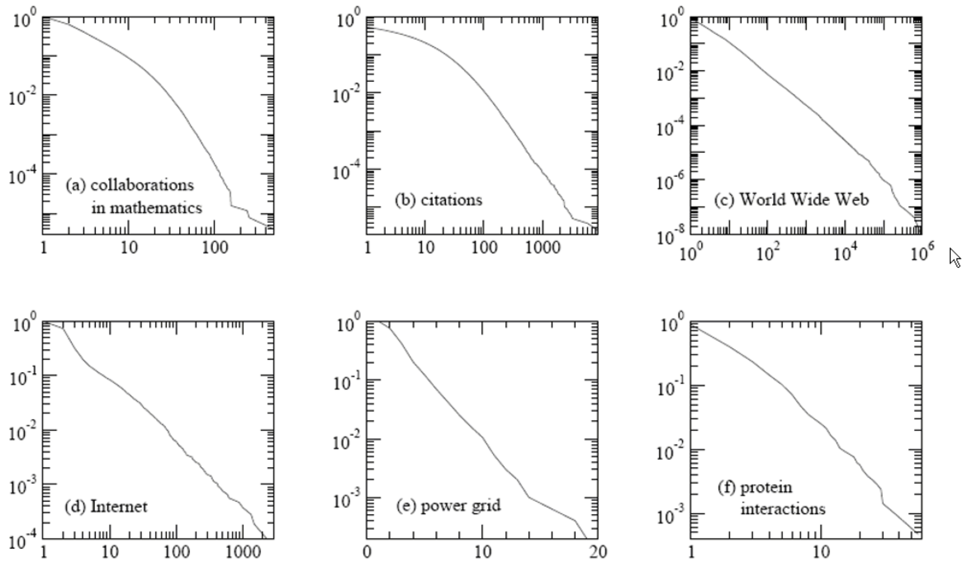
Cumulative distributions P_k of example real world NW



[1]

Degree Distribution

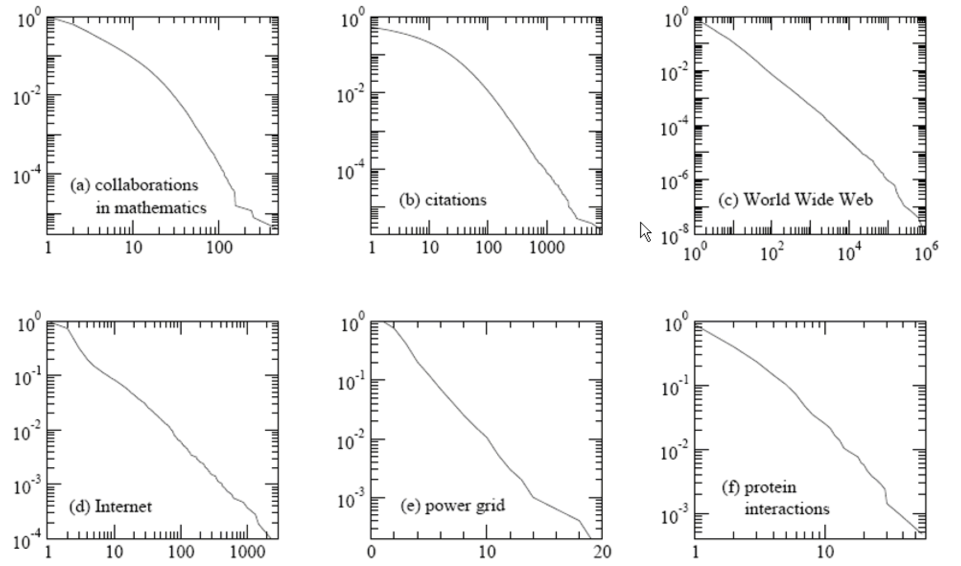
Cumulative distributions P_k of example real world NW



[1]

Degree Distribution

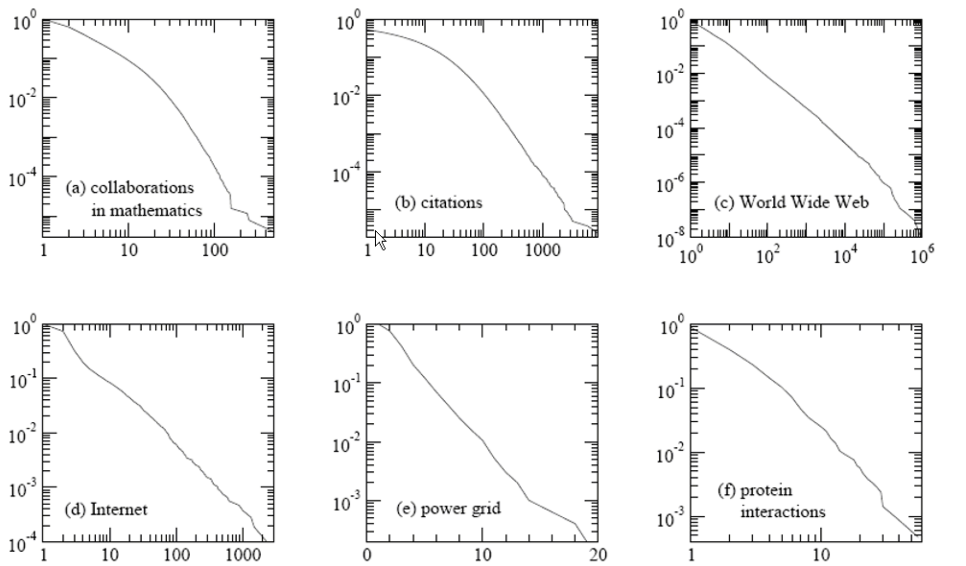
Cumulative distributions P_k of example real world NW



[1]

Degree Distribution

Cumulative distributions P_k of example real world NW



[1]

Degree Distribution

"Power law" == "Scale free":

- $f(x) = x^{\alpha}$ is only solution to functional equation formalizing scale freedom $f(ax) = b\, f(x)$
- in other words: change of scale → f still „looks the same"
- other point of view:

Although we can compute the expectation $E(k)=\Sigma_k\, k\, k^{-\alpha}$ if $\alpha>1$,

the variance (error bars) $Var(k)= \Sigma_k\, (k-E(k))^2\, k^{-\alpha}$

diverges → we „cannot be shure about k"

→ „no characteristic scale" → „scale free"

Degree Distribution

“Power law” == “Scale free”:

- $f(x) = x^{\alpha}$ is only solution to functional equation formalizing scale freedom $f(ax) = b\,f(x)$
- in other words: change of scale → f still „looks the same“
- other point of view:

Although we can compute the expectation $E(k)=\Sigma_k\, k\, k^{-\alpha}$ if $\alpha>1$,
the variance (error bars) $Var(k)= \Sigma_k\, (k-E(k))^2\, k^{-\alpha}$
diverges → we „cannot be shure about k“
→ „no characteristic scale“ → „scale free“

Degree Distribution

“Power law” == “Scale free”:

- $f(x) = x^{\alpha}$ is only solution to functional equation formalizing scale freedom $f(ax) = b\,f(x)$
- in other words: change of scale → f still „looks the same“
- other point of view:

Although we can compute the expectation $E(k)=\Sigma_k\, k\, k^{-\alpha}$ if $\alpha>1$,
the variance (error bars) $Var(k)= \Sigma_k\, (k-E(k))^2\, k^{-\alpha}$
diverges → we „cannot be shure about k“
→ „no characteristic scale“ → „scale free“