Gaussian Mixture Models

- Fuzzy C-Means is “OK” as a non-crisp clustering alg. but (as K-Means) favors spherical clusters → better approaches

Example: Gaussian Mixture Models (GMM)

- Linear combination of Gaussians
  \[ p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k) \quad \text{where} \quad \sum_{k=1}^{K} \pi_k = 1, \quad 0 \leq \pi_k \leq 1 \]

parameters to be estimated
Gaussian Mixture Models

- Fuzzy C-Means is “OK” as a non-crisp clustering alg, but (as K-Means) favors spherical clusters → better approaches

Example: Gaussian Mixture Models (GMM)

- Linear combination of Gaussians
  \[ p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k) \quad \text{where} \quad \sum_{k=1}^{K} \pi_k = 1, \ 0 \leq \pi_k \leq 1 \]

Machine Learning

- For a distribution \( p(x|\theta) \) parametrized by a set of parameters \( \theta \) and iid data \( X = \{x_1, x_2, ..., x_N\} \), simple machine learning corresponds to finding the \( \theta \) that best explains the data

  - iid: “identically independently drawn” \( \Rightarrow p(X|\theta) = \prod_i p(x_i|\theta) \)

  - \( p(X|\theta) \) is called likelihood

  - “finding the \( \theta \) that best explains the data”: Maximum Likelihood: \( \theta_{ML} = arg\max_\theta p(X|\theta) \Rightarrow \nabla_\theta p(X|\theta) = 0 \)

  - convenient: use \( \log p(X|\theta) \) instead of \( p(X|\theta) \)
    \( \Rightarrow \log p(X|\theta) = \sum_i \log p(x_i|\theta) \)
For a distribution \( p(x|\theta) \) parametrized by a set of parameters \( \theta \) and iid data \( X = \{x_1, x_2, ..., x_N\} \), simple machine learning corresponds to finding the \( \theta \) that best explains the data:

- iid: "identically independently drawn" \( \Rightarrow p(X|\theta) = \prod_i p(x_i|\theta) \)

\( p(X|\theta) \) is called likelihood

- "finding the \( \theta \) that best explains the data":
  - Maximum Likelihood: \( \theta_{ML} = \arg \max_\theta p(X|\theta) \Rightarrow \forall_\theta p(X|\theta) \neq 0 \)

  - convenient: use \( \log p(X|\theta) \) instead of \( p(X|\theta) \)
    \( \Rightarrow \log p(X|\theta) = \sum_i \log p(x_i|\theta) \)
Example: \( x \in \mathbb{R}^m \) and \( p(x|\theta) \) is one multivariate Gaussian

\[
p(x|\theta) = N(x|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2} \lvert \Sigma \rvert^{1/2}} \exp \left\{ -\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu) \right\}
\]

- log likelihood: \((\text{see base c})\)

\[
\ln p(X|\theta) = \ln p(X|\mu, \Sigma) = -\frac{ND}{2} \ln(2\pi) - \frac{N}{2} \ln \lvert \Sigma \rvert - \frac{1}{2} \sum_{n=1}^{N} (x_n - \mu)^T \Sigma^{-1} (x_n - \mu)
\]

- Maximum log likelihood:

\[
\theta_{ML} = \underset{\theta}{\operatorname{argmax}} \log p(X|\theta) \Rightarrow \mathbb{E}_\theta \left( \sum_i \log p(x_i|\theta) \right) = 0
\]

\[
\begin{align*}
\mu_{ML} : & \quad \frac{\partial}{\partial \mu} \ln p(X|\mu, \Sigma) = 0 \\
\Sigma_{ML} : & \quad \frac{\partial}{\partial \Sigma} \ln p(X|\mu, \Sigma) = 0
\end{align*}
\]

\[
\begin{align*}
\mu_{ML} &= \frac{1}{N} \sum_{n=1}^{N} x_n \\
\Sigma_{ML} &= \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu_{ML})(x_n - \mu_{ML})^T
\end{align*}
\]
Example: $x \in \mathbb{R}^m$ and $p(x|\theta)$ is one multivariate Gaussian

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

- log likelihood: (see base e)

$$\ln p(X|\theta) = \ln p(X|\mu, \Sigma) = -\frac{ND}{2} \ln (2\pi) - \frac{N}{2} \ln |\Sigma| - \frac{1}{2} \sum_{n=1}^{N} (x_n - \mu)^T \Sigma^{-1} (x_n - \mu)$$

- Maximum log likelihood:

$$\Theta_{ML} = \arg\max_{\theta} \ln p(X|\theta) = \nabla_{\theta} \left( \sum_i \ln p(x_i|\theta) \right) = 0$$

$$\mu_{ML} : \quad \frac{\partial}{\partial \mu} \ln p(X|\mu, \Sigma) = 0 \quad \Rightarrow \quad \mu_{ML} = \frac{1}{N} \sum_{n=1}^{N} x_n$$

$$\Sigma_{ML} : \quad \frac{\partial}{\partial \Sigma} \ln p(X|\mu, \Sigma) = 0 \quad \Rightarrow \quad \Sigma_{ML} = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu_{ML}) (x_n - \mu_{ML})^T$$

GMM: $p(x|\theta) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k), \quad 0 \leq \pi_k \leq 1, \quad \sum_{k=1}^{K} \pi_k = 1$

\[ 1 \]

\[ 0 \quad 0.5 \quad 1 \]

\[ 0 \quad 2.5 \]
GMM: \[ p(x|\theta) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k). \quad 0 \leq \pi_k \leq 1 \quad \sum_{k=1}^{K} \pi_k = 1 \]

• 1 of K representation

\[ z_k \in \{0, 1\} \text{ and } \sum_k z_k = 1 \]

\[ p(z_k = 1) = \pi_k \]

\[ p(z) = \prod_{k=1}^{K} \pi_k^{z_k} \]

• conditional probability

\[ p(x|z_k = 1) = N(x|\mu_k, \Sigma_k) \quad p(x|z) = \prod_{k=1}^{K} N(x|\mu_k, \Sigma_k)^{z_k} \]

\[ p(x) = \sum_{z} p(z)p(x|z) = \pi_k N(x|\mu_k, \Sigma_k) \]

\[ p(x,z) \]

GMM: \[ p(x|\theta) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k). \quad 0 \leq \pi_k \leq 1 \quad \sum_{k=1}^{K} \pi_k = 1 \]

• 1 of K representation

\[ z_k \in \{0, 1\} \text{ and } \sum_k z_k = 1 \]

\[ p(z_k = 1) = \pi_k \]

\[ p(z) = \prod_{k=1}^{K} \pi_k^{z_k} \]

• conditional probability

\[ p(x|z_k = 1) = N(x|\mu_k, \Sigma_k) \quad p(x|z) = \prod_{k=1}^{K} N(x|\mu_k, \Sigma_k)^{z_k} \]

\[ p(x) = \sum_{z} p(z)p(x|z) = \pi_k N(x|\mu_k, \Sigma_k) \]

\[ p(x,z) \]
GMM-Basics

\[ p(x|\theta) = p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k). \quad 0 \leq \pi_k \leq 1 \]
\[ \sum_{k=1}^{K} \pi_k = 1 \]

- 1 of K
  representation
  
  \( z_k \in \{0, 1\} \) and \( \sum_k z_k = 1 \)
  
  \[ p(z_k = 1) = \pi_k \]
  
  \[ p(z) = \prod_{k=1}^{K} \pi_k^{z_k} \]

- conditional
  probability

  \[ p(x|z_k = 1) = N(x|\mu_k, \Sigma_k) \]
  
  \[ p(x) = \prod_{k=1}^{K} N(x|\mu_k, \Sigma_k)^{z_k} \]

- Responsibilities

  \[ \gamma(z_k) \equiv p(z_k = 1|x) = \frac{p(z_k = 1)p(x|z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)p(x|z_j = 1)} = \frac{\pi_k N(x|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x|\mu_j, \Sigma_j)} \]

- Example

\[ \]

GMM-Basics

\[ p(x|\theta) = p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k). \quad 0 \leq \pi_k \leq 1 \]
\[ \sum_{k=1}^{K} \pi_k = 1 \]

- 1 of K
  representation
  
  \( z_k \in \{0, 1\} \) and \( \sum_k z_k = 1 \)
  
  \[ p(z_k = 1) = \pi_k \]
  
  \[ p(z) = \prod_{k=1}^{K} \pi_k^{z_k} \]

- conditional
  probability

  \[ p(x|z_k = 1) = N(x|\mu_k, \Sigma_k) \]
  
  \[ p(x|z) = \prod_{k=1}^{K} N(x|\mu_k, \Sigma_k)^{z_k} \]

- Responsibilities

  \[ \gamma(z_k) \equiv p(z_k = 1|x) = \frac{p(z_k = 1)p(x|z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)p(x|z_j = 1)} = \frac{\pi_k N(x|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x|\mu_j, \Sigma_j)} \]

- Example

\[ \]

1 of K
representation

\( z_k \in \{0, 1\} \) and \( \sum_k z_k = 1 \)

\[ p(z_k = 1) = \pi_k \]

\[ p(z) = \prod_{k=1}^{K} \pi_k^{z_k} \]

conditional
probability

\[ p(x|z_k = 1) = N(x|\mu_k, \Sigma_k) \]

\[ p(x|z) = \prod_{k=1}^{K} N(x|\mu_k, \Sigma_k)^{z_k} \]

Responsibilities

\[ \gamma(z_k) \equiv p(z_k = 1|x) = \frac{p(z_k = 1)p(x|z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)p(x|z_j = 1)} = \frac{\pi_k N(x|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j N(x|\mu_j, \Sigma_j)} \]

Example
**GMM-Basics**

- **Responsibilities**

  \[
  \gamma(z_k) \equiv p(z_k = 1|x) = \frac{p(z_k = 1)p(x|z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)p(x|z_j = 1)} = \frac{\pi_k \mathcal{N}(x|\mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x|\mu_j, \Sigma_j)}.
  \]

- **Example**

  ![Example](image)
GMM-Basics

- Responsibilities
  \[ \gamma(z_k) \equiv p(z_k = 1|x) = \frac{p(z_k = 1)p(x|z_k = 1)}{ \sum_{j=1}^{K} p(z_j = 1)p(x|z_j = 1) } = \frac{\pi_k N(x|\mu_k, \Sigma_k)}{ \sum_{j=1}^{K} \pi_j N(x|\mu_j, \Sigma_j) } \]

- Example

  ![Example](image)

  (a)  (b)  (c)

GMM-Basics

Maximum likelihood (GMM)

\[ \ln p(X|\pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k N(x_n|\mu_k, \Sigma_k) \right\} \]

- Vector of \( K \) D-dim. means \( \mu_k \)
- Vector of \( K \) DxD covariances \( \Sigma_k \)

- maximizing w.r.t. \( \pi, \mu \) and \( \Sigma \) \[
\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})x_n \\
\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k)(x_n - \mu_k)^T \\
\left( N_k = \sum_{n=1}^{N} \gamma(z_{nk}) \right) \pi_k = \frac{N_k}{N}
\]
GMM-Basics

Maximum likelihood (GMM)

\[
\ln p(X|\pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k) \right\}
\]

\* maximizing w.r.t. \(\pi, \mu\) and \(\Sigma\) \(
\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) x_n
\)
\(\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k)(x_n - \mu_k)^T
\)
\(N_k = \sum_{n=1}^{N} \gamma(z_{nk})\)
\(\pi_k = \frac{N_k}{N}\)

GMM-Basics

Maximum likelihood (GMM)

\[
\ln p(X|\pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k) \right\}
\]

\* maximizing w.r.t. \(\pi, \mu\) and \(\Sigma\) \(
\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) x_n
\)
\(\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k)(x_n - \mu_k)^T
\)
\(N_k = \sum_{n=1}^{N} \gamma(z_{nk})\)
\(\pi_k = \frac{N_k}{N}\)

GMM-Basics

Maximum likelihood (GMM)

\[
\ln p(X|\pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k) \right\}
\]

\* maximizing w.r.t. \(\mu\) and \(\Sigma\) \(
\mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) x_n
\)
\(\Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k)(x_n - \mu_k)^T
\)
\(N_k = \sum_{n=1}^{N} \gamma(z_{nk})\)
\(\pi_k = \frac{N_k}{N}\)

GMM-Basics

Maximum likelihood (GMM)

\[
\ln p(X|\pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k) \right\}
\]

\* maximizing w.r.t. \(\pi\) \(
\pi_k = \frac{\pi_k}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x|\mu_j, \Sigma_j)}
\)

GMM-Basics

Maximum likelihood (GMM)

\[
\ln p(X|\pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k) \right\}
\]

\* maximizing w.r.t. \(\pi\) \(
\pi_k = \frac{\pi_k}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x|\mu_j, \Sigma_j)}
\)

GMM-Basics

Maximum likelihood (GMM)

\[
\ln p(X|\pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k) \right\}
\]

\* maximizing w.r.t. \(\pi\) \(
\pi_k = \frac{\pi_k}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x|\mu_j, \Sigma_j)}
\)

Example
Maximum likelihood (GMM)

\[ \ln p(X|\pi, \mu, \Sigma) = \ln \left( \sum_{n=1}^{N} \prod_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k) \right) \]

Vector of \( K \) \( D \)-dim. means \( \mu_k \)
Vector of \( K \) \( D \times D \) covariances \( \Sigma_k \)

- maximizing w.r.t \( \pi, \mu \) and \( \Sigma \)

\[ \mu_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})x_n \]
\[ \Sigma_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k)(x_n - \mu_k)^T \]
\[ N_k = \sum_{n=1}^{N} \gamma(z_{nk}) \]
\[ \pi_k = \frac{N_k}{N} \]

so what?! \( \Rightarrow \) Problem: Expr. depend on \( \gamma(z_{nk}) \) which depends on \( \pi, \mu, \Sigma \)
which depends on \( \gamma(z_{nk}) \) which depends on …..

Idea: Alternating approach (EM-algorithm):

Step t: Evaluate \( \gamma(z_{nk}) \) using \( (\pi, \mu, \Sigma)_{(t-1)} \)
Evaluate \( (\pi, \mu, \Sigma)_{(t)} \) using \( \gamma(z_{nk})_{(t)} \)
- Having latent variables $Z$, ML becomes

$$\ln p(X|\theta) = \ln \sum_{Z} p(X, Z|\theta)$$

- Summation inside $\ln \rightarrow$ Problems!

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

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

- alternating EM:

E-Step: compute

$$Q(\theta, \theta^{old}) = \sum_{Z} p(Z|X, \theta^{old}) \ln p(X, Z|\theta).$$

M-Step: compute

$$\theta^{new} = \arg \max_{\theta} Q(\theta, \theta^{old}).$$

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

$$p(x|\mu_k, \Sigma_k) = \frac{1}{(2\pi)^{1/2}} \exp \left\{ -\frac{1}{2\epsilon} ||x - \mu_k||^2 \right\}$$

- we get for the responsibilities:

$$\gamma(z_{nk}) = \frac{\pi_k \exp \left\{ -||x_n - \mu_k||^2/2\epsilon \right\}}{\sum_j \pi_j \exp \left\{ -||x_n - \mu_j||^2/2\epsilon \right\}}$$

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

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

$\rightarrow$ same as on slide 18
**EM: Relation to K-Means**

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

  $$p(x|\mu_k, \Sigma_k) = \frac{1}{(2\pi)^{d/2}} \exp \left\{ -\frac{1}{2\epsilon} \|x - \mu_k\|^2 \right\}$$

- we get for the responsibilities:

  $$\gamma(z_{nk}) = \frac{\pi_k \exp \left\{ -\frac{1}{2\epsilon} \|x_n - \mu_k\|^2 \right\}}{\sum_j \pi_j \exp \left\{ -\frac{1}{2\epsilon} \|x_n - \mu_j\|^2 \right\}}$$

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

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

  $\to$ same as on slide 18

---

**EM: Relation to K-Means**

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

  $$p(x|\mu_k, \Sigma_k) = \frac{1}{(2\pi)^{d/2}} \exp \left\{ -\frac{1}{2\epsilon} \|x - \mu_k\|^2 \right\}$$

- we get for the responsibilities:

  $$\gamma(z_{nk}) = \frac{\pi_k \exp \left\{ -\frac{1}{2\epsilon} \|x_n - \mu_k\|^2 \right\}}{\sum_j \pi_j \exp \left\{ -\frac{1}{2\epsilon} \|x_n - \mu_j\|^2 \right\}}$$

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

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

  $\to$ same as on slide 18

---

**Studying Complex Networks**

- Paradigm shift: small NW $\to$ large NW:
  - interest in individual elements ("centrality of node x") $\to$
  - interest in global statistical / topological properties ("degree distribution of NW")
  - investigating particular NW instance $\to$
  - general model for types of NW with certain properties
  - 100 nodes $\to$ $10^8$ nodes
  - visualization possible $\to$ impossible / pointless

- Typical sorts of NW investigated:
  - social, information, technological, biological
Studying Complex Networks

- Paradigm shift: small NW → large NW:
  - interest in individual elements ("centrality of node x") →
    interest in global statistical / topological properties ("degree
    distribution of NW")
  - investigating particular NW instance →
    general model for types of NW with certain properties
  - 100 nodes → 10^8 nodes
  - visualization possible → impossible / pointless

- Typical sorts of NW investigated:
  social, information, technological, biological

Social Networks

- Formalizations of social context (mostly long and medium term)

  - Until dawn of databases: Collection via social science methods
    (questionnaires etc.) → „laboratory effects“;
  - Until dawn of Social Web: Indirect characterization of social
    relations (e.g. co-citation networks, collaboration NW (e.g. movie
    co-acting NW))
  - Today (Social Web / Mobile Social Web): self declared
    explication of social structures; „collect data about“ / „observe“ Homo Sapiens in its „natural habitat“ (→ Twitter,
    Facebook etc.)

Information Networks / Knowledge NW

- Most studied examples: citation NW (tree), the WWW;

- Example findings:
  - p(k) of author having k papers: p(k) ~ k^-a : power law
  - distribution of in or out degrees of WWW pages (also for citation
    NW): p(k) ~ k^-a

- Other examples:
  - bipartite preference networks :
    → recommender systems == link prediction on these NW;
    example: collaborative filtering
  - ontologies, semantic NW
  - word networks
  - tripartite tag/author/item networks
    → Folksonomies

Information Networks / Knowledge NW

- Most studied examples: citation NW (tree), the WWW;

- Example findings:
  - p(k) of author having k papers: p(k) ~ k^-a : power law
  - distribution of in or out degrees of WWW pages (also for citation
    NW): p(k) ~ k^-a

- Other examples:
  - bipartite preference networks :
    → recommender systems == link prediction on these NW;
    example: collaborative filtering
  - ontologies, semantic NW
  - word networks
  - tripartite tag/author/item networks
    → Folksonomies
Information Networks / Knowledge Networks

- **Most studied examples:** citation NW (tree), the WWW;

- **Example findings:**
  - $p(k)$ of author having $k$ papers: $p(k) \sim k^{-\alpha}$: power law
  - distribution of in or out degrees of WWW pages (also for citation NW): $p(k) \sim k^{-\alpha}$

- **Other examples:**
  - bipartite preference networks:
    - $\Rightarrow$ recommender systems $\Rightarrow$ link prediction on these NW;
      example: collaborative filtering
  - ontologies, semantic NW
  - word networks
  - tripartite tag/author/item networks
    - $\Rightarrow$ Folksonomies

Technological Networks

- **Most studied examples:** distribution NW:
  - the Internet,
  - electric power grids,
  - traffic NW (roads, railway tracks etc.)

- **Biological / Chemistry / Physics Networks**

  - **Most studied examples:**
    - biochemical pathways, gene-protein and protein-protein interaction NW
    - nervous systems, vascular systems (also natural distribution NW),
    - food NW, ecological dependency NW
Technological Networks

- Most studied examples: distribution NW:
  - the Internet,
  - electric power grids,
  - traffic NW (roads, railway tracks etc.))

Biological / Chemistry / Physics Networks

- Most studied examples:
  - biochemical pathways, gene-protein and protein-protein interaction NW
  - nervous systems, vascular systems (also natural distribution NW),
  - food NW, ecological dependency NW

FIG. 2. Three examples of the kinds of networks that are the topic of this review. (a) A food web of predator-prey interactions between species in a freshwater lake [77]. Picture courtesy of Nao Tanizawa and Richard Williams. (b) The network of collaborations between scientists at a private research institution [114]. (c) A network of sexual contacts between individuals in the study by Pentz et al. [142].