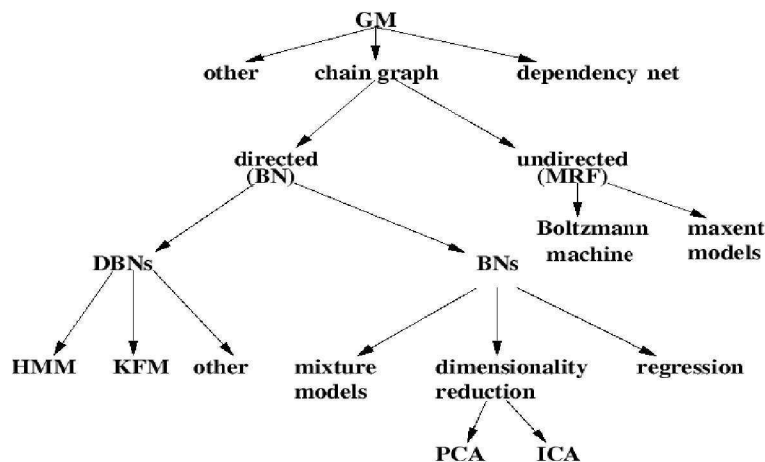


## Graphical Models (GM)

Graphical models are a marriage between probability theory and graph theory. They provide a natural tool for dealing with two problems that occur throughout applied mathematics and engineering - uncertainty and complexity - and in particular they are playing an increasing important role in the design and analysis of machine learning algorithms. Fundamental to the idea of a graphical model is the notion of modularity - a complex system is built by combining simpler parts. Probability theory provides the glue whereby the parts are combined, ensuring that the system as a whole is consistent, and providing ways to interface models to data. The graph theoretic side of graphical models provides both an intuitively appealing interface by which humans can model highly-interacting sets of variables as well as a data structure that lends itself naturally to the design of efficient general-purpose algorithms.

The graphical model formalism generalizes many probabilistic approaches and models in fields such as statistics, systems engineering, information theory, pattern recognition, and statistical mechanics. Some examples are mixture models, factor analysis, hidden Markov models, Kalman filters, Bayesian networks, Boltzmann machines, and the Ising model. The framework of graphical model provides a way to represent these systems as instances of a common formalism.

There are two main types of graphical models: undirected and directed graphical models (see Figure 1).



**Fig. 1:** Taxonomy of GM.

Undirected graphical models (also known as Markov Random Fields) are popular in the physics and vision communities. Directed graphical models (also known as Bayesian networks, belief networks,...) are widespread in the artificial intelligence and machine

learning communities. Also a combination of directed and undirected graphs is available. This form of graph is called chain graph. Factor graphs fit into the framework of graphical models. These graphs are used for coding. Also the approximate inference algorithm called loopy belief propagation used for Graphical Models is well known for turbo codes.

There are several topics what might be interesting to cover in the course:

- Introduction to Graphical Models (GM): Taxonomy of GM, directed/undirected GM, factor Graphs, Dynamic Bayesian Networks, Independence, Conditional Independence, d-separation, Bayes Ball Algorithm, Examples,...
- Parameter Learning (ML, MAP) and Structure Learning (Structural EM, K2,...)
- Exact Inference: Variable Elimination, Junction Tree algorithm (Message passing, Belief propagation, sum product algorithm,)
- Loopy Belief Propagation (sum product algorithm revisited, LBP)
- Generalized Belief Propagation Algorithms
- Variational Inference
- Sampling: Gibbs Sampling, MCMC

Applications:

- Factor graphs (Coding (Turbo codes,...))
- Kalman Filter/Linear Gaussian Models, HMM,
- Bayesian network classifiers (discriminative/generative Parameter Learning)
- GM in Speech Recognition/Language Modeling
- Particle Filter
- Bioinformatics

Literature (Literature is not restricted to the following suggestions):

#### **General Literature - Books:**

- F. V. Jensen. "**Bayesian Networks and Decision Graphs**". Springer. 2001.  
Probably the best introductory book available.
- R. G. Cowell, A. P. Dawid, S. L. Lauritzen and D. J. Spiegelhalter. "**Probabilistic Networks and Expert Systems**". Springer-Verlag. 1999.  
Probably the best book available, although the treatment is restricted to exact inference.
- M. I. Jordan (ed). "**Learning in Graphical Models**". MIT Press. 1998.  
Loose collection of papers on machine learning, many related to graphical models.  
One of the few books to discuss *approximate* inference.

- B. Frey. "Graphical models for machine learning and digital communication", MIT Press. 1998. Discusses pattern recognition and turbocodes using (directed) graphical models.
- F. Jensen. "An introduction to Bayesian Networks". UCL Press. 1996. Out of print. Superseded by his 2001 book.
- S. Lauritzen. "Graphical Models", Oxford. 1996.  
The definitive mathematical exposition of the theory of graphical models. (Very tough)
- J. Pearl. "Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference." Morgan Kaufmann. 1988.  
The book that got it all started! A very insightful book, still relevant today.

### **Learning**

- Buntine Wray, A Guide to the Literature on Learning Probabilistic Networks From Data. IEEE transactions On Knowledge and Data Engineering 1996
- Paul J. Krause Learning Probabilistic Networks. 1998
- N. Friedman, The Bayesian Structural EM Algorithm., UAI, 1998.
- G.F. Cooper and E. Herskovits, "A Bayesian Method for the Induction of Probabilistic Networks from Data", Machine Learning, 9, pp. 309-347, 1992.
- Heckerman, D. A tutorial on learning with Bayesian networks. In Learning in Graphical Models, M. Jordan, ed.. MIT Press, Cambridge, MA, 1999

### **Learning with hidden variables: the EM algorithm**

- R.M. Neal and G. Hinton "A View Of The Em Algorithm That Justifies Incremental, Sparse, And Other Variants", Learning in Graphical Models, Kluwer, 1998
- A. O. Dempster, N. M. Laird, and D. B. Rubin, Maximum likelihood from incomplete data via the em algorithm., Journal of the Royal Statistical Society, Series B (Methodological), vol. 39, no. 1, pp. 1--38, 1977.
- Bilmes, J. A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. Technical Report, University of Berkeley, ICSI-TR-97-021, 1997

### **Exact Inference:**

- Huang, C., and Darwinche, A. Inference in belief networks: a procedural guide. International Journal of Approximate Reasoning, Vol 15, no 3, pp 225-263, 1996
- different books form above

### **Generalized Belief Propagation Algorithms:**

- Yedidia, J.S.; Freeman, W.T.; Weiss, Y., "Understanding Belief Propagation and Its Generalizations", TR-2001-22, 2002.
- Yedidia, J.S.; Freeman, W.T.; Weiss, Y., , "Constructing Free Energy Approximations and Generalized Belief Propagation Algorithms", TR-2004-40, 2004.

### **Variational Inference:**

- Jordan, M. I., Ghahramani Z., Jaakkola T. S. Saul Lawrence K. An introduction to variational methods for graphical models. In Learning in Graphical Models, M. Jordan ed. MIT Press, Cambridge, MA 1999
- T. Jaakkola "Tutorial on variational approximation methods", 2000.
- T. Jaakkola and M. Jordan, 1998. "Variational probabilistic inference and the QMR-DT database"
- Z. Ghahramani and M.J. Jordan, "Factorial Hidden Markov Models", Machine Learning, 1997.
- Z. Ghahramani and G.E. Hinton, "Variational Learning for Switching State-Space Models", Neural Computation (?),

### **Loopy Belief Propagation**

- Y. Weiss. Correctness of local probability propagation in graphical models with loops. Neural Computation, 2000.
- K. P. Murphy, Y. Weiss, and M. Jordan. Loopy belief propagation for approximate inference: an empirical study. In Proceedings of Uncertainty in AI, pages 467-475, 1999.
- Y. Weiss and W. Freeman. Correctness of belief propagation in gaussian graphical models of arbitrary topology. In NIPS, volume 12, 1999.

- S.M. Aji, G.B. Horn, and R.J. McEliece. On the convergence of iterative decoding on graphs with a single cycle. In Proc. 1998 ISIT, 1998.
- R. J. McEliece, D. J. MacKay and J. F. Cheng. Turbo decoding as an instance of Pearl's belief propagation algorithm, IEEE Journal of Selected Areas of Communication, pp. 140--152, Feb, 1998.
- J. S. Yedidia, W. T. Freeman and Y. Weiss. Bethe free energies, Kikuchi approximations, and belief propagation algorithms. TR 2001-16, 2001.

### **Sampling: Gibbs, MCMC**

- MacKay D., An introduction to MonteCarlo methods. In Learning in Graphical Models, M. Jordan, ed.. MIT Press, Cambridge, MA, 1999
- Koller, Lerner, Angelov, A General Algorithm for Approximate Inference and Its Application to Hybrid Bayes Nets (1999),
- Introduction to Monte Carlo Methods, D.J.C. MacKay, Learning in Graphical Models, M.I. Jordan, Kluwer Academic Publishers, 1998

### **HMMs:**

- Rabiner, L.R. A tutorial on hidden Markov models and selected applications in speech recognition, Volume 77, Issue 2, Feb 1989 Pages 257-286 (A standard HMM reference and tutorial)
- Fine S, Singer Y and Tishby N. The hierarchical hidden Markov model: Analysis and Applications. Machine Learning, 32(1), July 1998.

### **Kalman Filter / Linear Gaussian Models**

- Sam Roweis & Zoubin Ghahramani, 1999. A Unifying Review of Linear Gaussian Models, Neural Computation 11(2) (1999) pp.305-345

### **Dynamic Bayesian Networks**

- Murphy K., Dynamic Bayesian networks: Representation, Inference and Learning, PhD Thesis, UC Berkeley, 2002.
- Ghahramani Zoubin, Learning Dynamic Bayesian networks, Oct. 1997

## **Factor Graphs**

- Kschischang, F.R. Frey, B.J., Loeliger, H.A. “Factor Graphs and the Sum- Product Algorithm”, IEEE Transactions on Information Theory, Vol. 47, No.2, 2001.

## **Bayesian Network Classifiers:**

- N. Friedman, D. Geiger, and M. Goldszmidt, “ Bayesian Network Classifiers”, Machine Learning, 1997.
- D. Grossman and P. Domingos, “Learning Bayesian Network Classifiers by Maximizing Conditional Likelihood”, ICML, 2004.
- R. Greiner, W. Zhou, X. Su, and B. Shen, “Structural Extension to Logistic Regression: Discriminative Parameter Learning of Belief Net Classifiers”, TR, 2004.

## **Markov Random Fields:**

- S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE-PAMI, 6, 1984, 721-741.
- Kindermann, R., and Snell, J. L. Markov Random Fields and their applications, American Mathematical Society, 1980