Signal processing for graphs

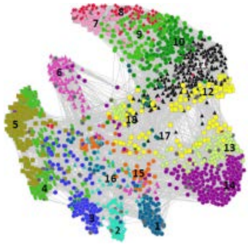
Alfred Hero

Communications and Signal Processing Seminar

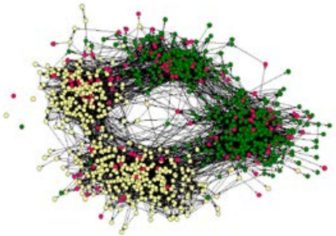
Sept 11, 2014

1. Network representations of data
2. Statistical summarization of random graphs
3. Probabilistic models of random graphs
4. Wrapup and conclusions
5. References

The Internet
(Burch and Cheswick, 1998)



Gene pathways
(Huang, 2011)



School friendships
(Moody, 2001)

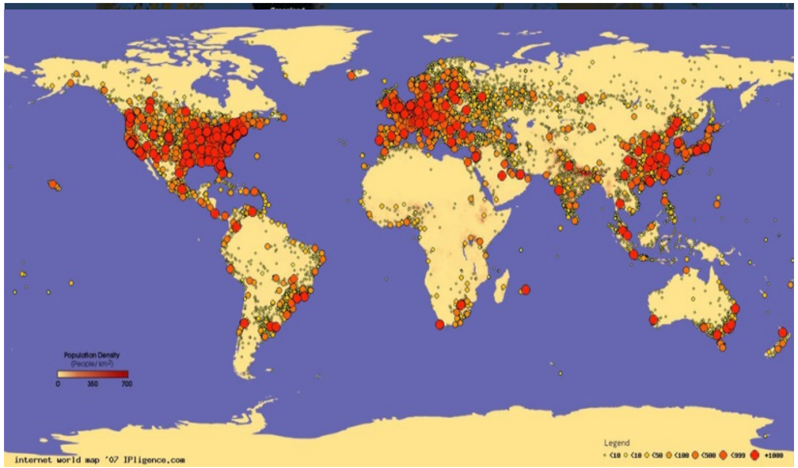
What this talk will not cover

- “Signal processing on graphs” – graph topology is SP substrate
  - Graphs as a finite field for SP algorithms like FFT, wavelets, clustering, spectral decomposition [Shuman, Narang, Frossard, Ortega, Vandergheynst 2013]
  - Distributed SP over graph [Dimakis, Kar, Moura, Rabbat, Scaglione 2010]
  - Distributed graphical models [Wiesel, H 2009], [Meng, Wei, Wiesel, H 2013]
- ”Signal processing with graphs”- graph used to estimate something else
  - Entropic graph estimators of entropy [H, Ma, Michel, Gorman 2001]
  - Chain and anti-chain graphs for information retrieval [Calder, Esedoglu, H, 2013]
- “Signal processing in graphs” – in situ probing of a physical network
  - Network tomography [Coates, H, Nowak, Yu 2002], [Shih, H, 2006]
  - Network probing for resiliency [Chen, H 2013, 2014]

Nor will we cover in any detail:

- Rendering of graphs or graph visualization [Xu, Kliger, H 2013]
- Multigraph models [Oselio, Kulesza, H 2013, 2014]
- Directed graph models [Wainright&Jordan 2008], [Rao, States, Engel, H 2007]
- Dynamic graph models [Westveld, Hoff 2011], [Xu, Kliger, H 2014]
- Phase transitions [Nadakuditi&Newman 2012], [Firouzi, H 2014], [Chen, H 2013]

I. Network representations of data

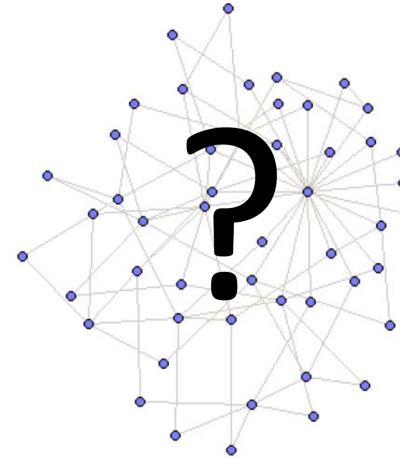


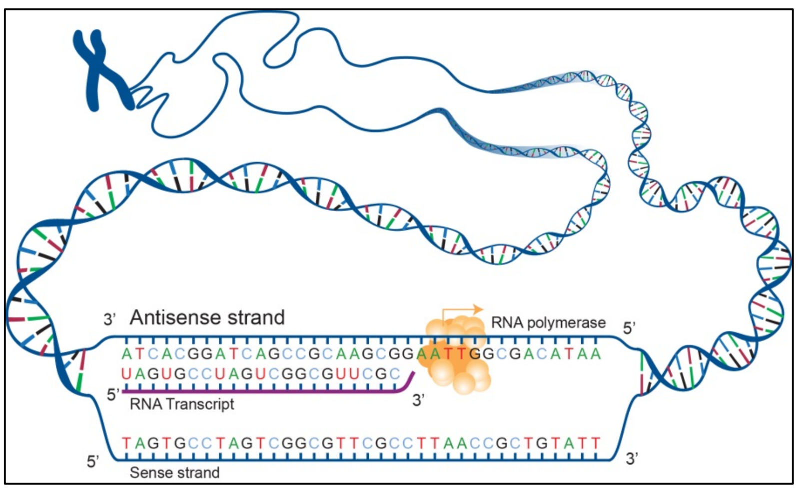
Spambot map

Honeypot traps
spam TCP-IP traces

Spammer Network

Observed interaction
Profile correlation



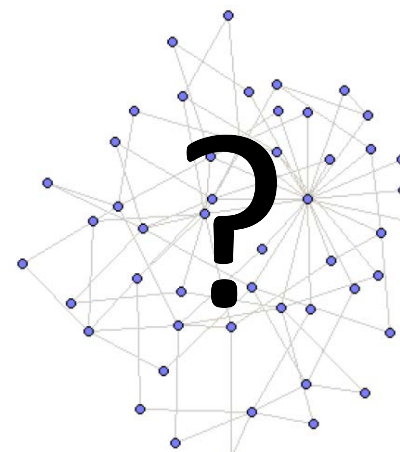


mRNA transcript map along DNA strand (NHGRI-85265)

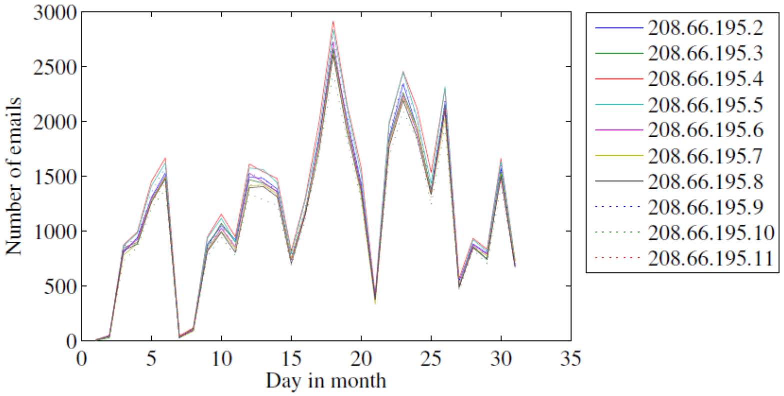
HiC sequencing
mRNA genechips

Genomic Network

Observed interaction
Profile correlation



Profile correlation

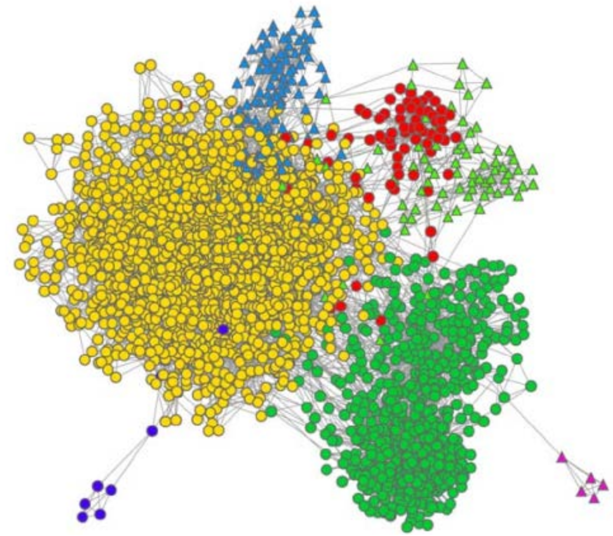


208.66.195/24 group of ten harvesters

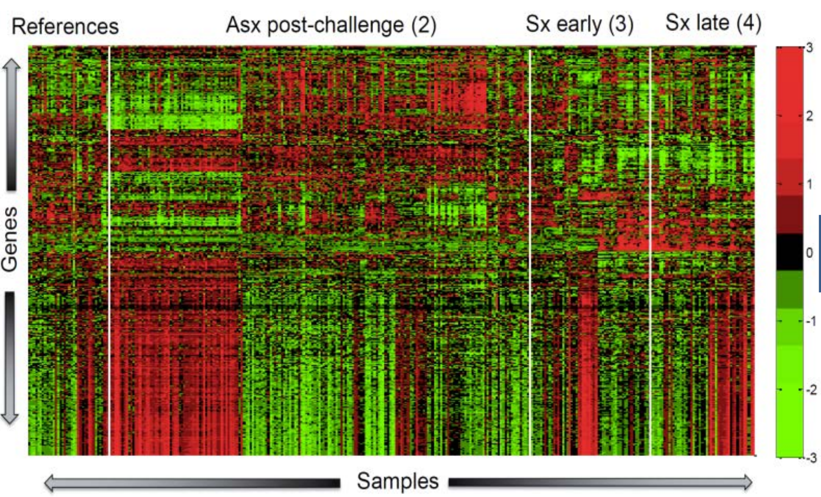
Honeypot traps spam TCP-IP traces

Spammer Network

Profile correlation



Xu, Kliger and H, ICC 2009

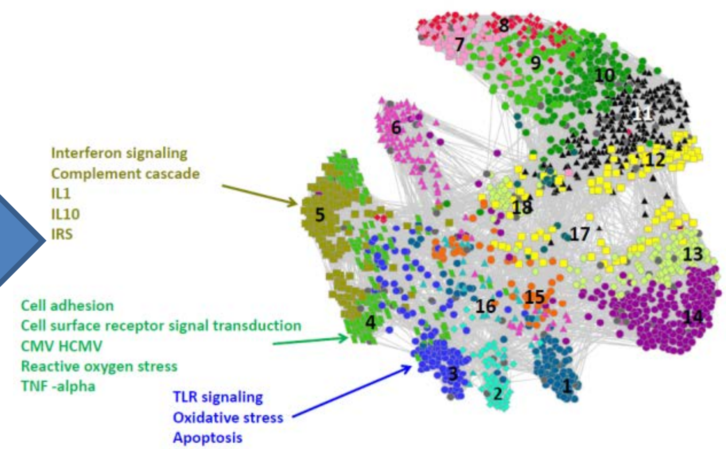


mRNA gene expression profiles

mRNA genechips

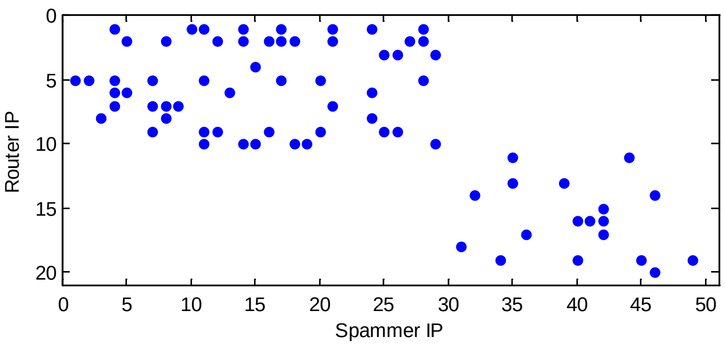
Genomic Network

Profile correlation



Huang et al, PLoS Genetics 2011

Observed interaction

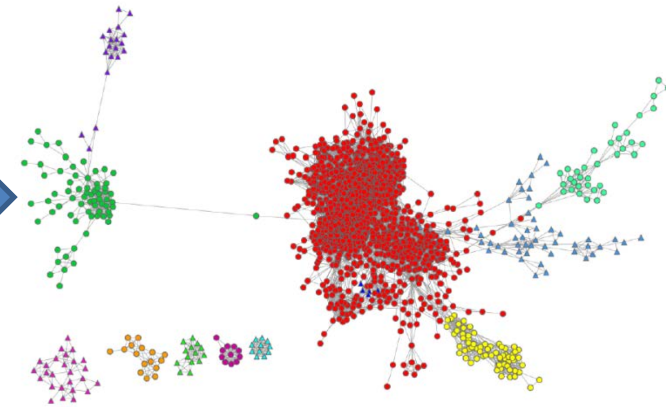


Email router usage matrix

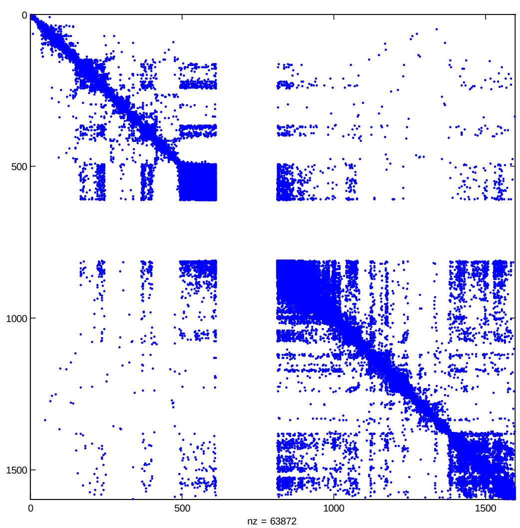
Honeypot traps
spam TCP-IP traces

Spammer Network

Observed interaction



Xu, Kliger and H, ICC 2009

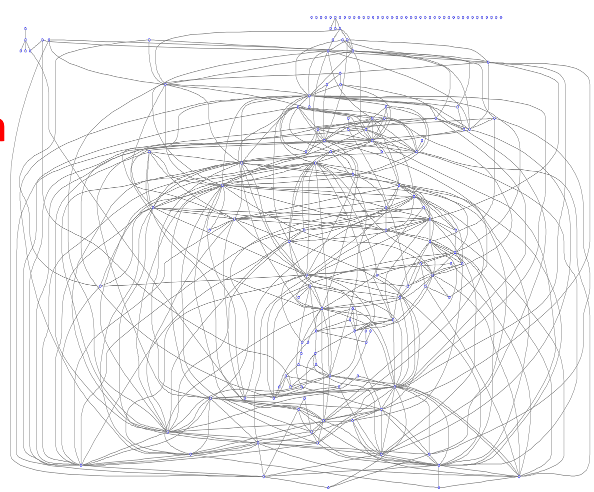


HiC contact matrix for chromosome 11

HiC – Genome-wide chromatin conformation capture

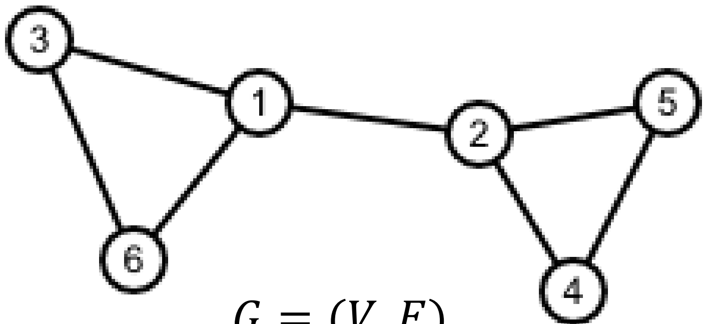
Genomic Network

Observed interaction



Chen, Comment, Chen, Ronquist, Ried, H, Rajapakse 2014

Graphs and adjacency/weight matrices



$G = (V, E)$

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

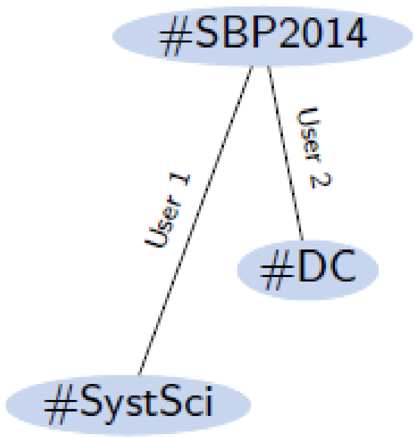
- A graph with n nodes is denoted $G = (V, E) \in \Omega_n, \quad |\Omega_n| = 2^{\binom{n}{2}} \approx 2^{2^n}$
  - V are vertices (nodes)
  - E are edges (links)
- In example on left:
  - $V = \{v_1, \dots, v_6\}$
  - $E = \{e_{12}, e_{13}, e_{16}, e_{24}, e_{25,} e_{36}, e_{45}\}$
- Vertices and edges can have attributes and weights, resp.

- The location/weight of edges in a graph are given by the adjacency matrix $\boldsymbol{A}$
- **Relational graphs**: edges ($\boldsymbol{A}$) are directly observed.
- **Associational (Behavioral) graphs**: edges are derived from node attributes

Attributional vs Relational data

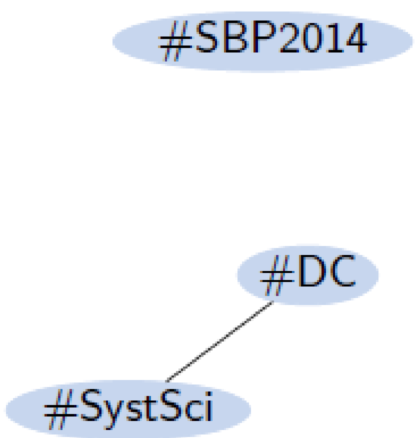
- Two broad categories of data for developing a model [Oselio, K, H 2013]
  - **Attributional**: node adjacencies estimated from node measurements
    - Node attributes are observed random variables. Edges reflects behavioral similarity of node attributes (similarity models, behavioral models)
      - Markov random fields, boolean networks: edges are latent variables
      - Examples: similar email semantics, Twitter #hashtag use, Facebook postings, tech content in publications, tastes in music
  - **Relational**: node adjacencies estimated from edge measurements
    - Edges or edge weights are observed random variables. An edge reflects a relation between node pair (familiarity models, coordination models)
      - Erdos-Renyi, exponential graphs: edge realizations observed
      - Examples: email exchange, Twitter follower, Facebook friend, co-authorship, biological relations
- In either case, the edges can be weighted or unweighted.
  - Weighted: edge strength encoded as edge length, color, thickness.
  - Unweighted: edge is binary - either present or absent

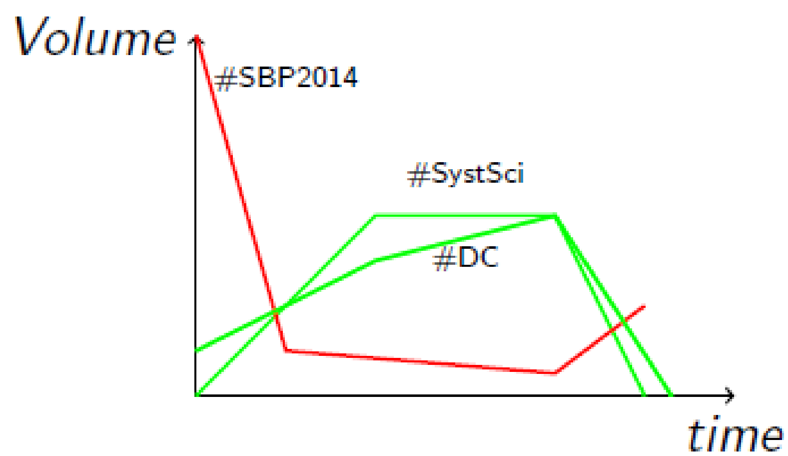
Example: Twitter hashtag multigraph



User 1: #SBP2014, #SystSci

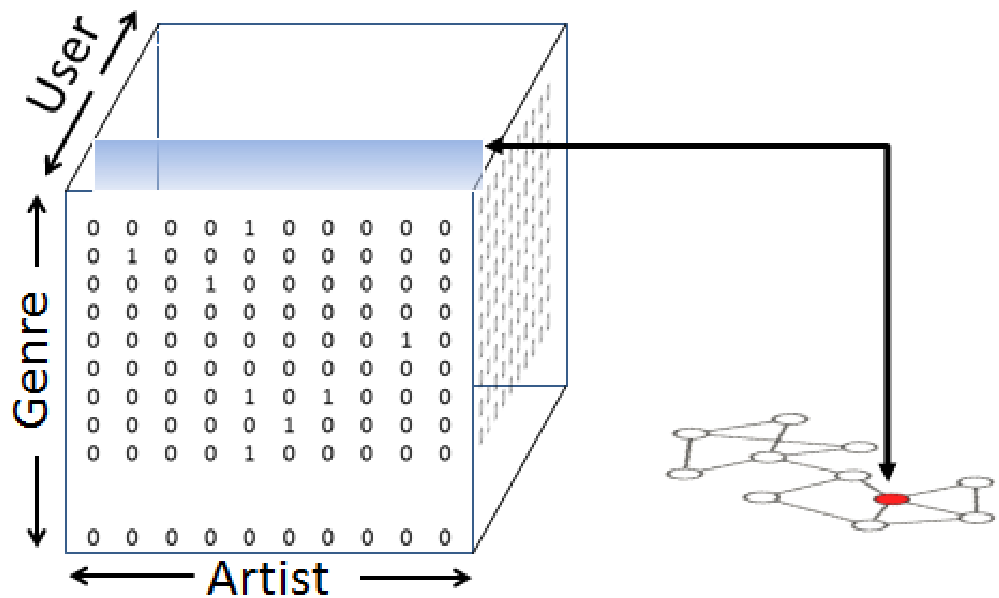
User 2: #SBP2014, #DC





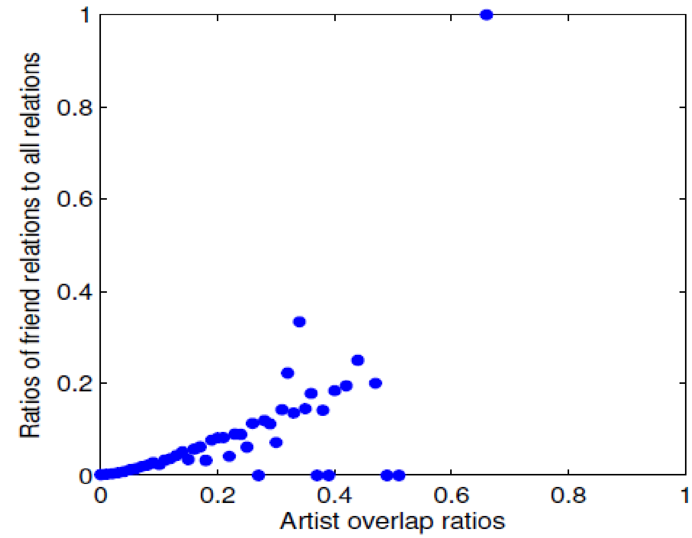
Oselio, Kulesza, H SBP 2014

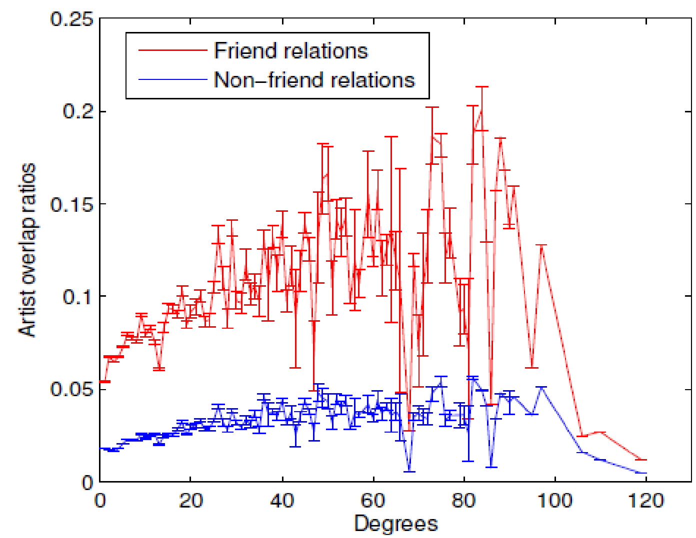
Example: Social collaborative retrieval



User preference matrix
(Attributional)

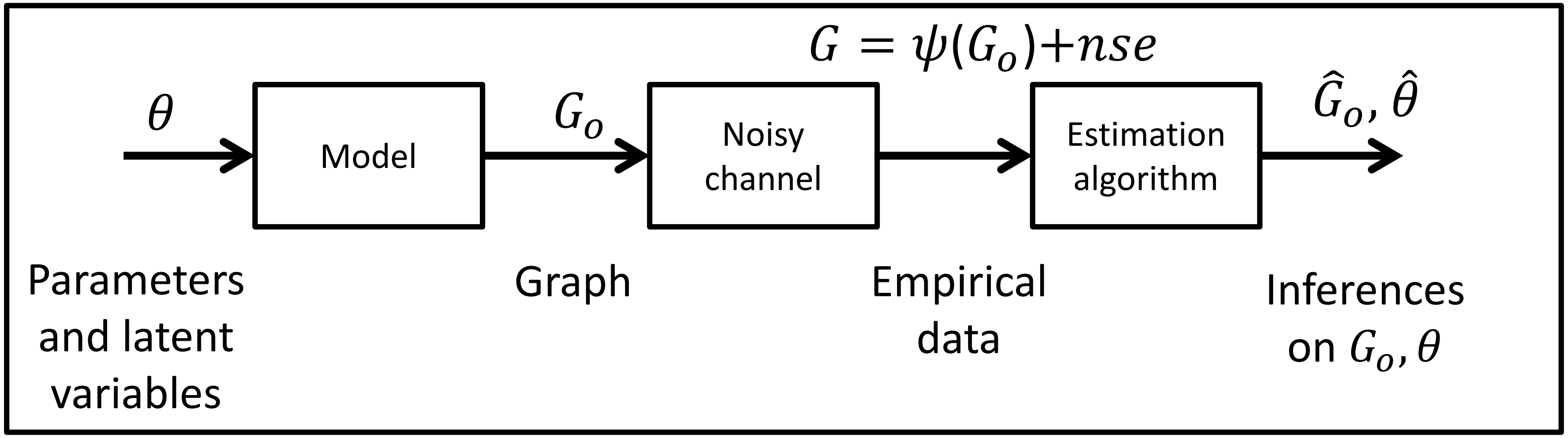
User social network
(Relational)





Hsiao, Kulesza, H 2014

The graph inference model



Focus here is modeling – summarization vs generative models

- Summarization (statistical) model
  - Graph summarized by a few statistics (on degree, centrality, paths)
  - Highly scalable for high dimensional graphs (many nodes, edges, states)
- Generative (probabilistic) model
  - Full probability distribution of graph is modeled (jpdf of nodes/edges)
  - Can suffer from poor scalability for high dimensional graphs

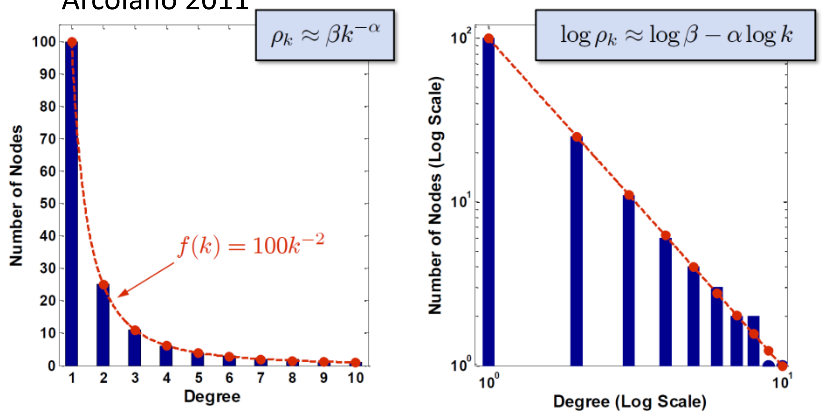
II. Summarization: Path statistics

- $\Pi = \{path(i,j)\}_{i>j}$: set of n(n-1)/2 shortest paths between all pairs of nodes in graph
- Diameter of graph: max(length(Π))
- Mean path length: average(length(Π))
- "Small-world" behavior [Strogatz and Watts 1998]
  - Smaller mean path length than would be found in a random graph
  - Lots more clustering than in a random graph: many triads (triangles)
  - "Nodes densely connected with few intermediaries" [Cho&Fowler 2010]
- Example: HEP co-authorship network [Newman 2001]
  - 8361 authors: 19,085 connections: mean path length=6.9
  - Compare with Poisson graph of same size: mean path length=24.4

Summarization: degree distribution

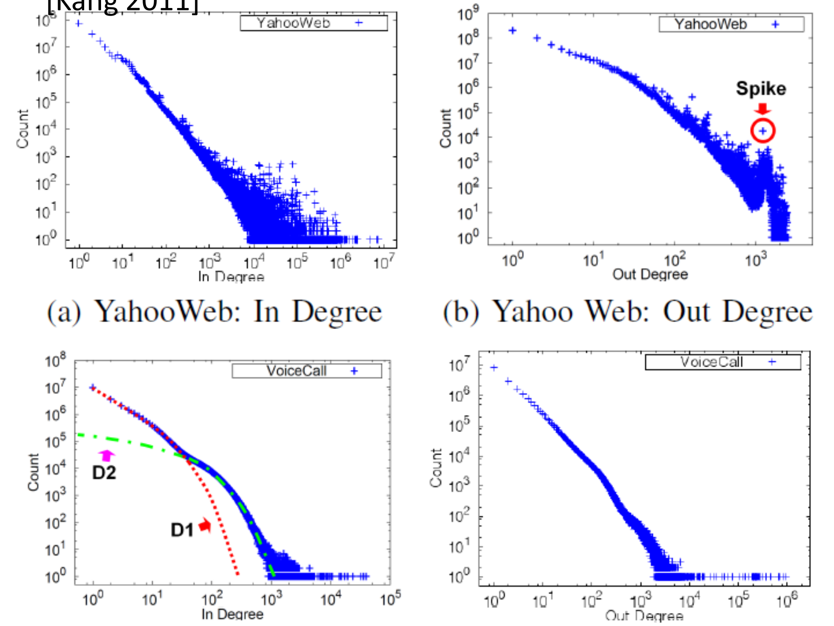
- Degree sequence: $\{d_1, \ldots, d_n\}$, $d_i = \sum_{j=1}^{n} a_{ij}$
- Degree histogram: $\rho_k = \sum_{i=1}^{n} I(d_i = k)$
- Power law model often proposed for real-world network data [Strogatz 1999, Newman 2001]

Arcolano 2011



Power law degree distribution

[Kang 2011]



(a) YahooWeb: In Degree (b) Yahoo Web: Out Degree

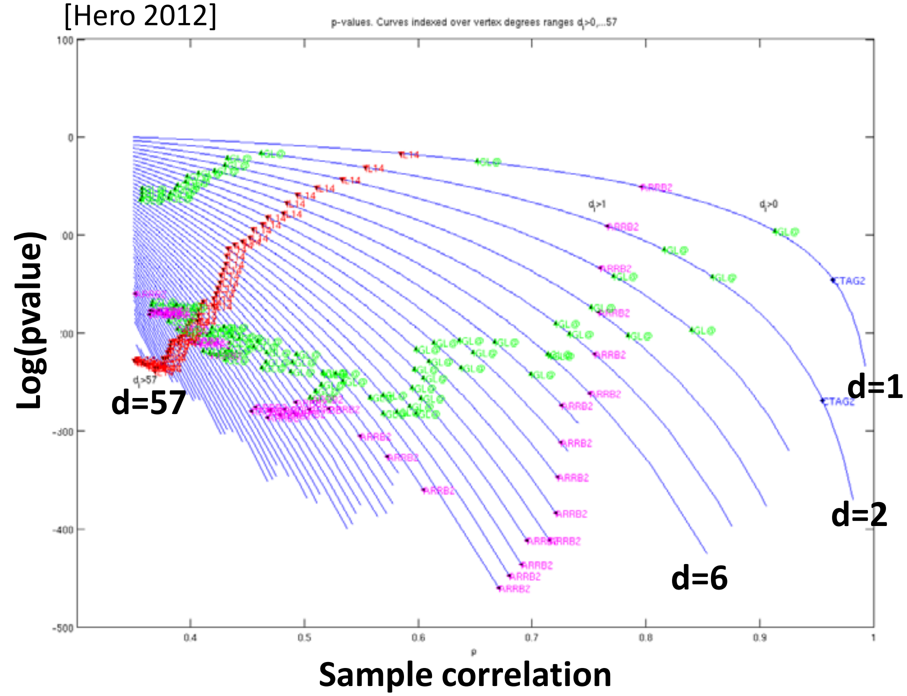
(e) VoiceCall: In Degree (f) VoiceCall: Out Degree

(a) Power law is good fit (low d)
(b) Log-normal is good fit (except high d)
(e) Log-normal mixture is good fit
(f) Power law is good fit

Summarization: p-value waterfall plot

- Introduced for attributional (correlation and partial correlation) graphs [H Rajaratnam 2012]
- Plot of p-values of each connected node as function of sample correlation or partial correlation.
- p-value=P(degree≥d|block sparse)
- Summing the number of nodes over each degree branch gives the degree histogram.
- Can be used to detect highly significant nodes in a large correlation graph

[Hero 2012]

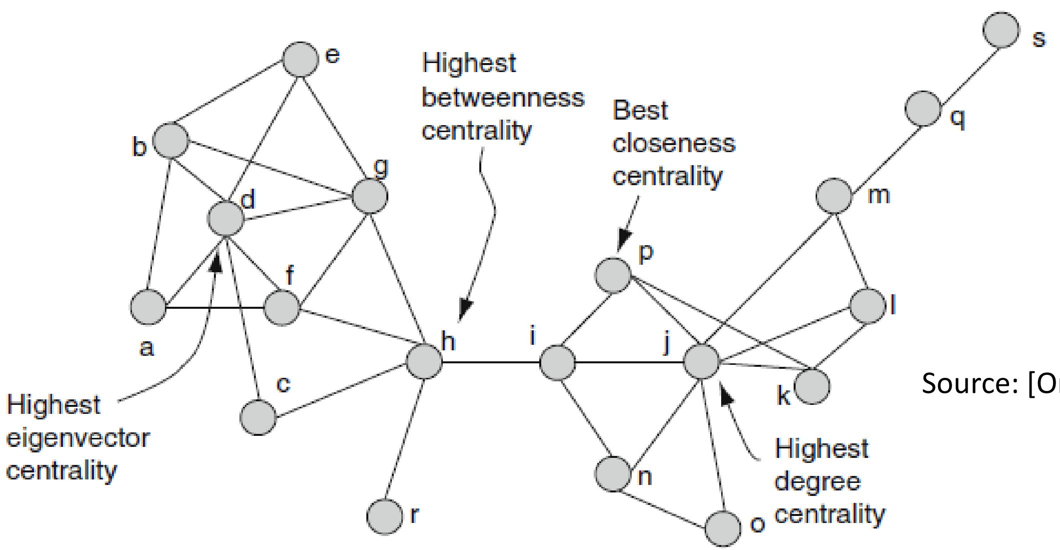


Waterfall plot for NKI Breast cancer data

- 24,481 nodes (Affy HU133 genes)
- 295 gene chips (used for sample corr)

Summarization: network centrality measures

- Network centrality captures the relative importance of a node to the global topology of the graph.
  - A node with high centrality is a "key player" in the network [Ortiz-Arroyo 2010]
  - Removal of a highly central node could severely disrupt the network
- Social network examples of high centrality nodes:
  - A popular individual - lots of friends in the network (**degree centrality**)
  - An individual with lowest average hop-distance from others (**closeness centrality**)
  - An individual who is sole liaison between two communities (**betweenness centrality**)
  - An individual who is popular among popular people (**eigenvector centrality**)



Source: [Ortiz-Arroyo 2010]

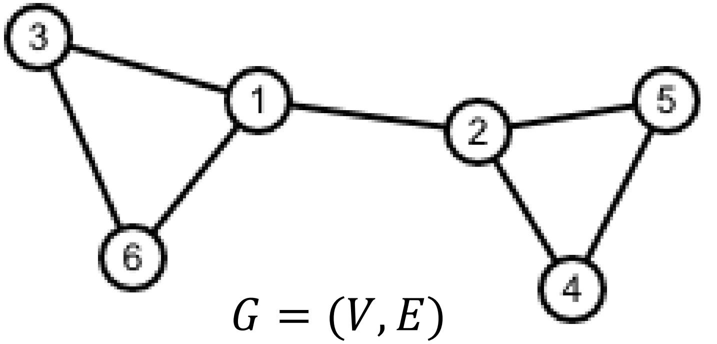
Degree centrality

- Degree centrality is a locally computable measure of centrality (C=O(n) computations).
- Number of direct connections to the node (vertex degree)

$$v_i = \sum_{j=1}^{n} a_{ij} \quad \Leftrightarrow \quad \boldsymbol{v} = \boldsymbol{A1}$$

- Examples:
  - Social network: social popularity of person *i* in a friendship network
  - Citation network: number of documents that cite *i*-th document

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$



$G = (V, E)$

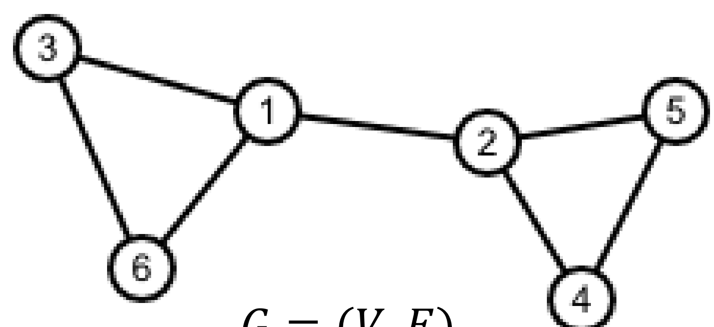
Closeness centrality

- Let H be the matrix of hop-distances (shortest path distance) between pairs of nodes (C=O($n^2 \log(n)$) computations).
- Closeness centrality measures avg closeness to other nodes

$$v_i = \left(\frac{1}{n} \sum_{j=1}^{n} h_{ij} \right)^{-1}$$

- Examples:
  - Social network: highly central person has low avg degree-of-separation
  - Citation network: Paul Erdos and Mark Newman have high centrality in mathematics and network science, respectively.

$$\boldsymbol{H} = \begin{pmatrix} 0 & 1 & 1 & 2 & 2 & 1 \\ 1 & 0 & 2 & 1 & 1 & 2 \\ 1 & 2 & 0 & 3 & 3 & 1 \\ 2 & 1 & 3 & 0 & 1 & 3 \\ 2 & 1 & 3 & 1 & 0 & 3 \\ 1 & 2 & 1 & 3 & 3 & 0 \end{pmatrix}$$



$G = (V, E)$

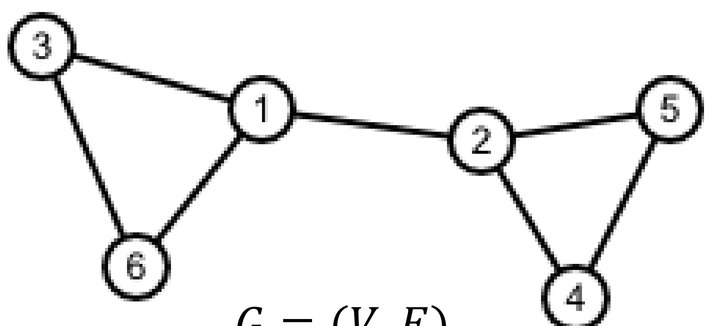
Betweenness centrality

- Average number of shortest-paths that pass through node *i*

$$v_i = \left(n(n-1)\right)^{-1} \sum_{j=1}^{n} \sum_{k=1}^{n} I(i \in path_{jk})$$

- Important nodes connect many other nodes (C=O(($n^2 \log(n)$)))
- Examples:
  - Social network: person who is critical link between large communities
  - Citation network: Author who publishes across very different disciplines

$$\boldsymbol{H} = \begin{pmatrix} 0 & 1 & 1 & 2 & 2 & 1 \\ 1 & 0 & 2 & 1 & 1 & 2 \\ 1 & 2 & 0 & 3 & 3 & 1 \\ 2 & 1 & 3 & 0 & 1 & 3 \\ 2 & 1 & 3 & 1 & 0 & 3 \\ 1 & 2 & 1 & 3 & 3 & 0 \end{pmatrix}$$



$G = (V, E)$

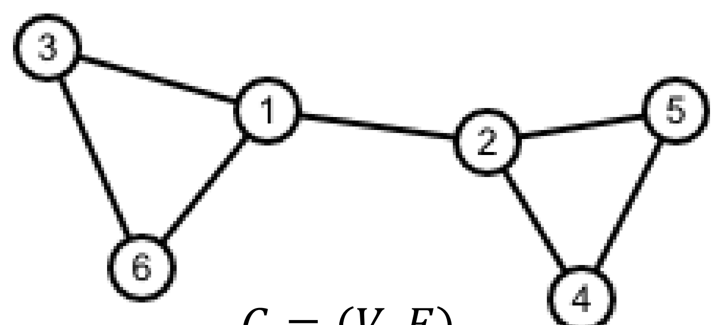
Eigenvector centrality

- A weighted measure of adjacency where centrality of *i*-th node is proportional to that of its neighbors

$$v_i \propto \sum_{j=1}^{n} a_{ij} v_j \qquad \Leftrightarrow \qquad \lambda \boldsymbol{v} = \boldsymbol{A}\boldsymbol{v}$$

- Solutions are eigenvectors and eigenvalues of of $\boldsymbol{A}$ (C=O(n^2))
- Centrality vector is $\boldsymbol{v}_1 =$ eigenvector associated with $\max(\lambda_j)$
- Examples:
  - Social network: popular individual among a popular group of friends
  - Citation network: paper that is highly cited by highly cited peers

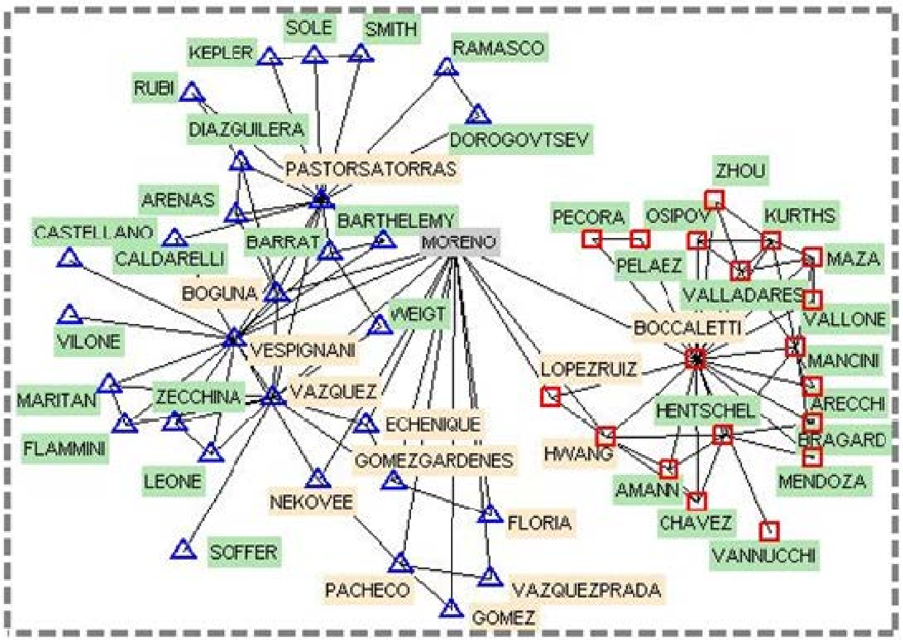
$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$



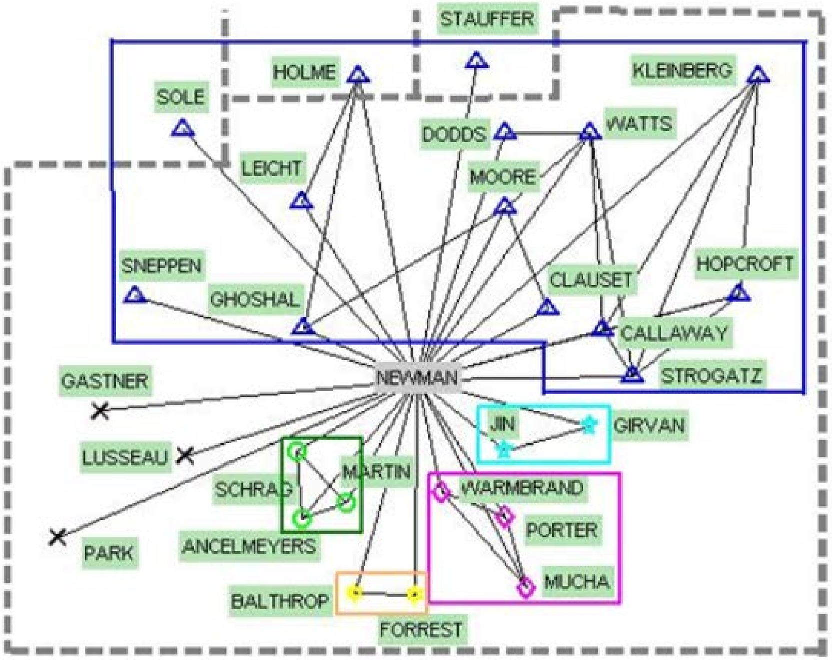
$G = (V, E)$

A centrality measure for finding polyglots

- **Local Fiedler Vector Centrality** [Chen, H 2013]: degree to which removal of a node from graph reduces algebraic connectivity
  - Algebraic connectivity: smallest number of node removals that disconnect graph
  - Fiedler vector **y** is 2^{nd} smallest eigenvector of **L**=**A**-diag(sum(**A**)) [Fiedler 1973]
  - **y'Ly** is a lower bound on the algebraic connectivity
  - LFVC of node i is: LFVC(i) = $\sum_{j \neq i} \left(y_i - y_j\right)^2$

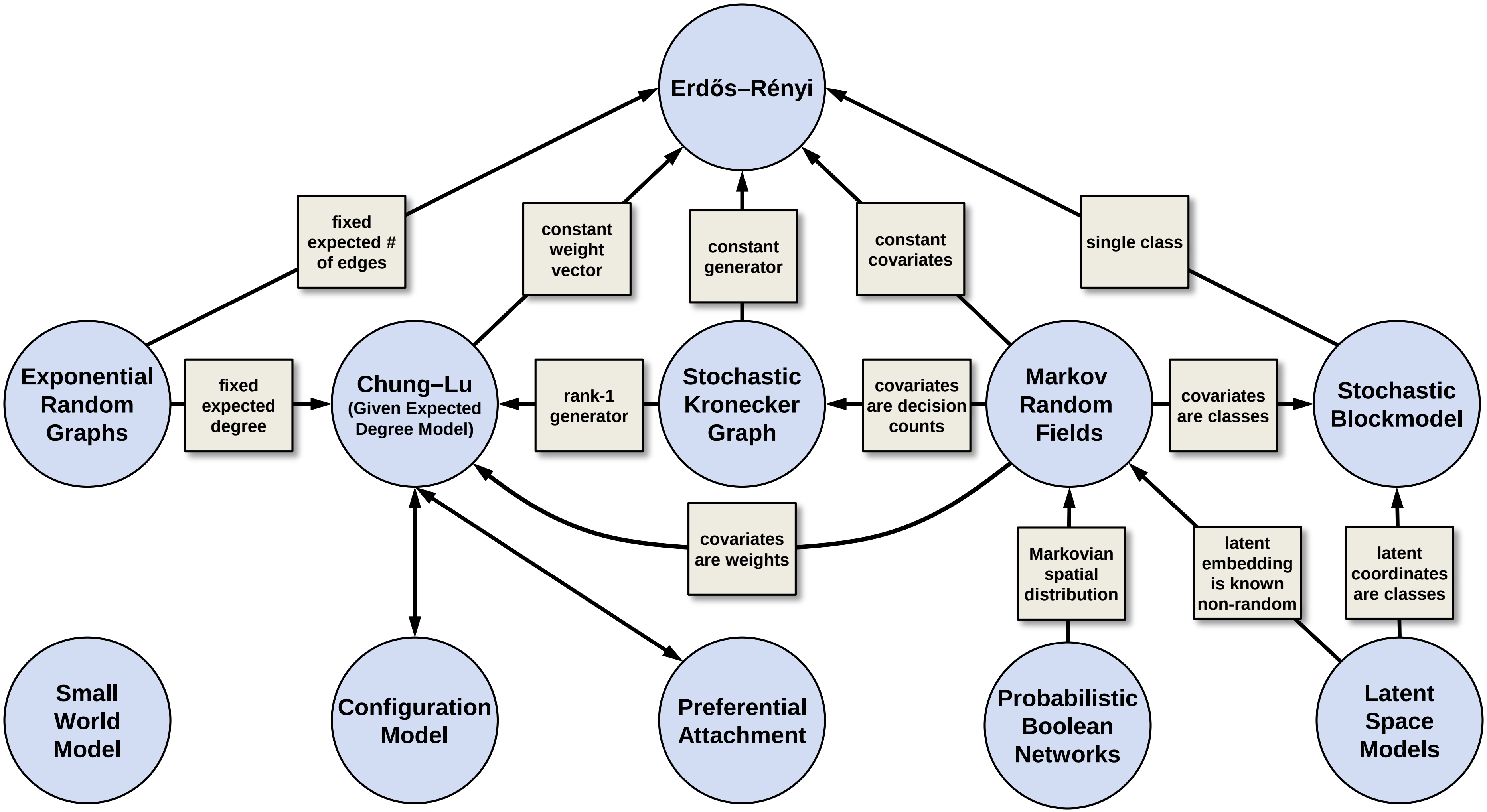


Yamir Moreno has top LFVC since connects two large communities in the network science co-author net.



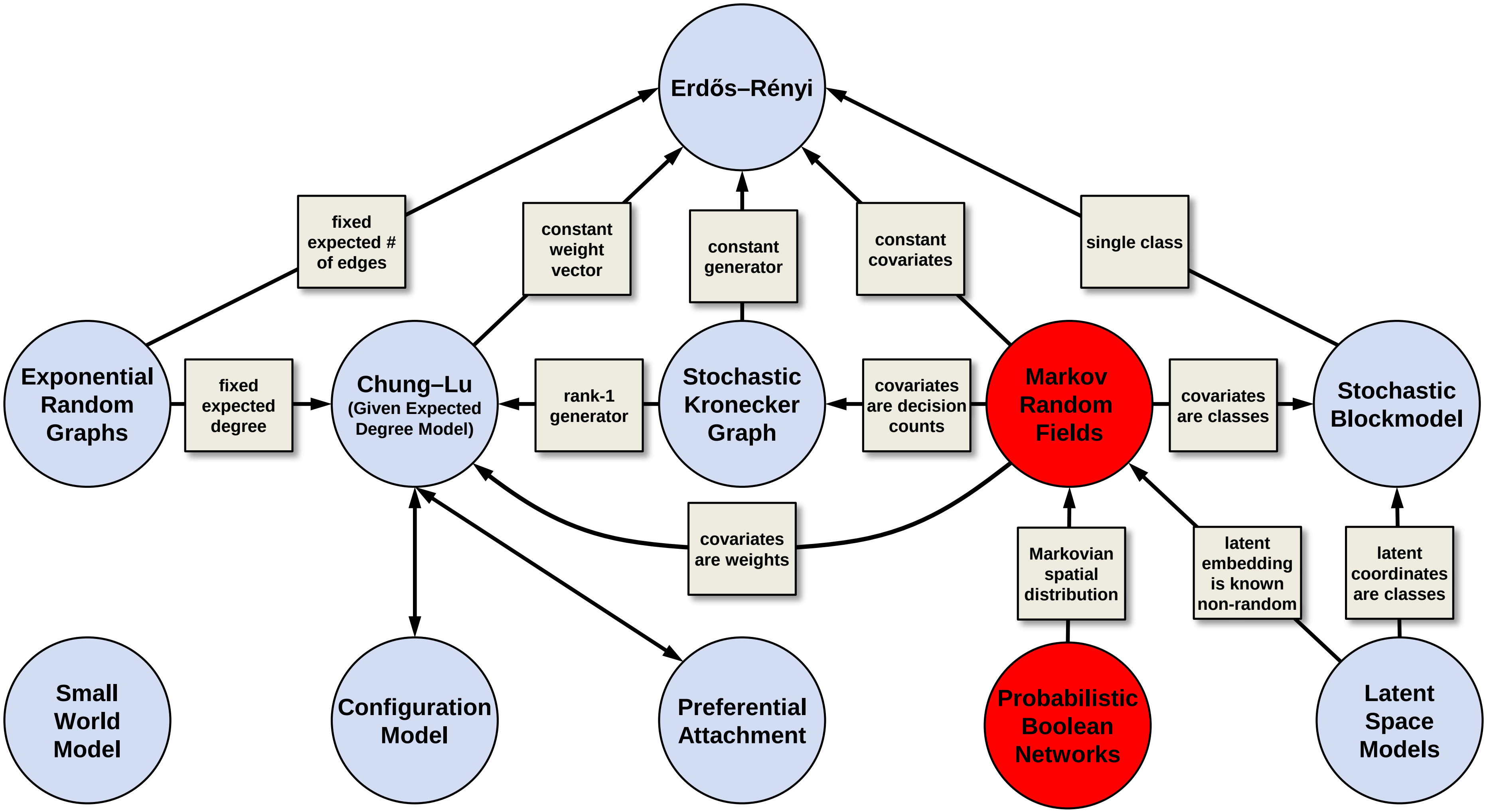
Mark Newman has 2^{nd} largest LFVC after Moreno is removed from network

III. Generative random graph models



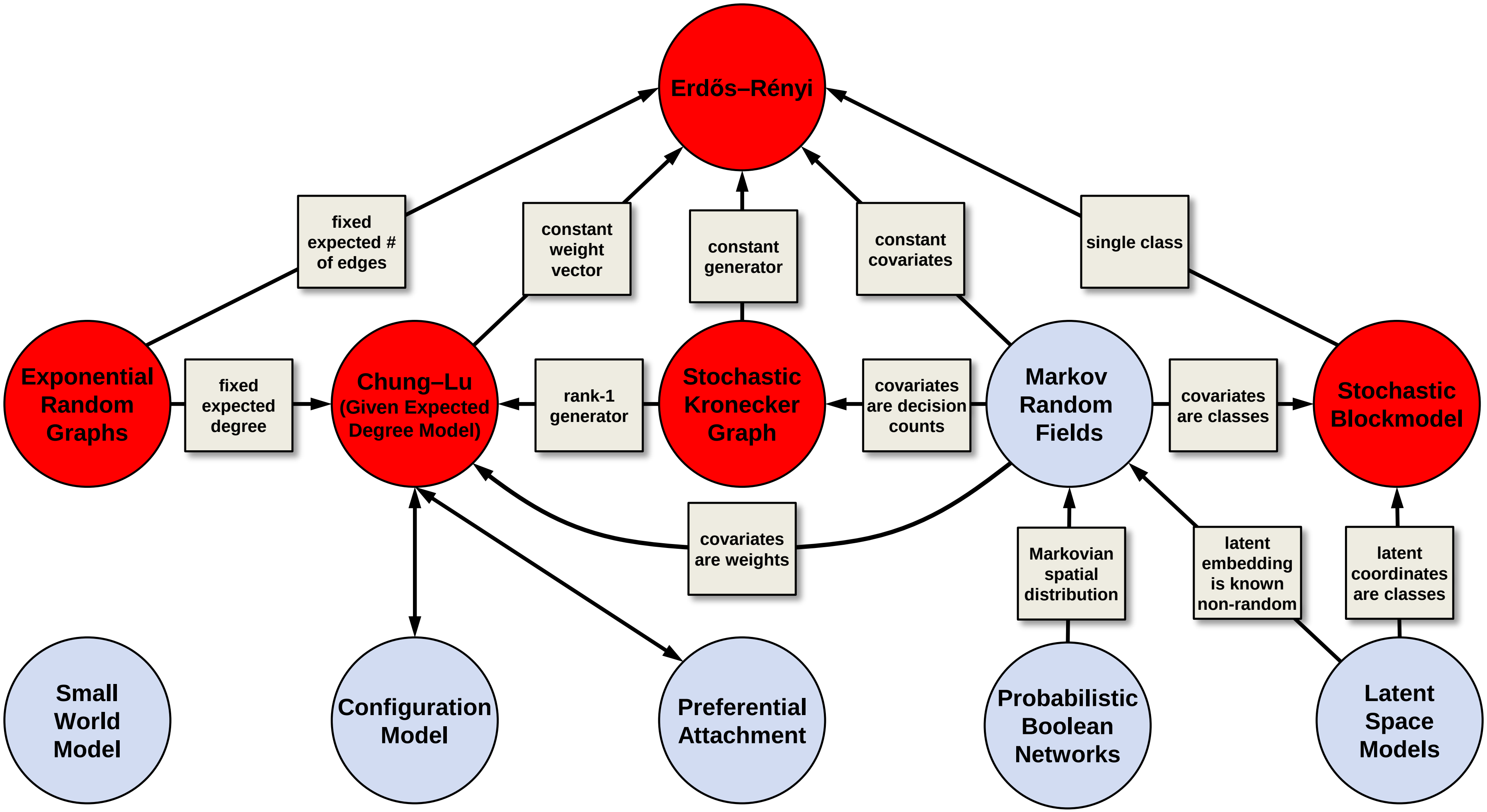
Adapted from Nicholas Arcolano "Statistical models for random graphs" 2011.

Random graph models for attributional data



Adapted from Nicholas Arcolano "Statistical models for random graphs" 2011.

Random graph models for relational data



Adapted from Nicholas Arcolano "Statistical models for random graphs" 2011.

Generative random graph models

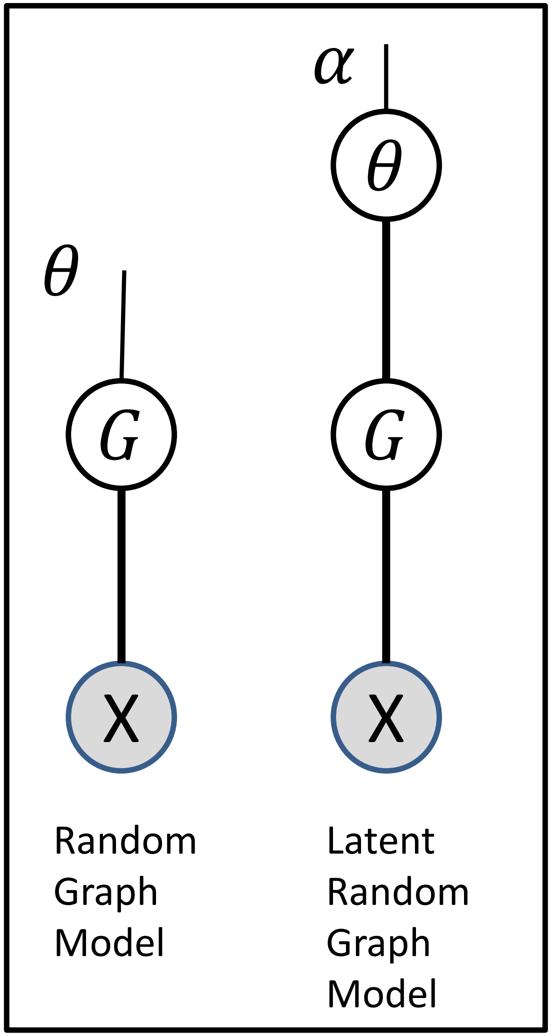
- Assume a prior distribution $p_G(G)$ on $G \in \Omega_p$
- Define conditional distribution p
  - Induces posterior distribution on G

$$p_{G|X}(G|X) = p_{X|G}(X|G)p_G(G)/p_X(X)$$

- Random graph model: $p_{X|G}$ and p_G depend on a fixed number of non-random parameters θ
- Latent random graph model: θ is random with pdf depending on additional parameter α
  - Markov property=conditional independence

$$p_{Z|G,\theta} = p_{Z|G}$$

- Bayesian inference of G is performed by fitting posterior to data
  - MAP or minMSE estimates of G, e.g., by MCMC, Belief Propagation (BP), or Laplace-Bernstein
  - Likelihood ratio test (LRT) of hypotheses on G



Refs: Kolaczyk 2009, Koller 2009, Wainright 2008.

Factor graph representations

- Let $\{\pi_j\}$ be N subsets of node indices ranging over $1, \ldots, n$
- Let $\{\eta_j\}$ be N subsets of edge indices ranging over $1, \ldots, \frac{n(n-1)}{2} = P$
- **Attributional** data: factor graph model for joint distribution of **node** attributes

$$p(x_1, \ldots, x_n) = \prod_{j=1}^{N} f_j(x_{\pi_j})$$

- **Relational** data: factor graph model for joint distribution of **edge** attributes

$$p(e_1, \ldots, e_P) = \prod_{j=1}^{N} f_j(e_{\eta_j})$$

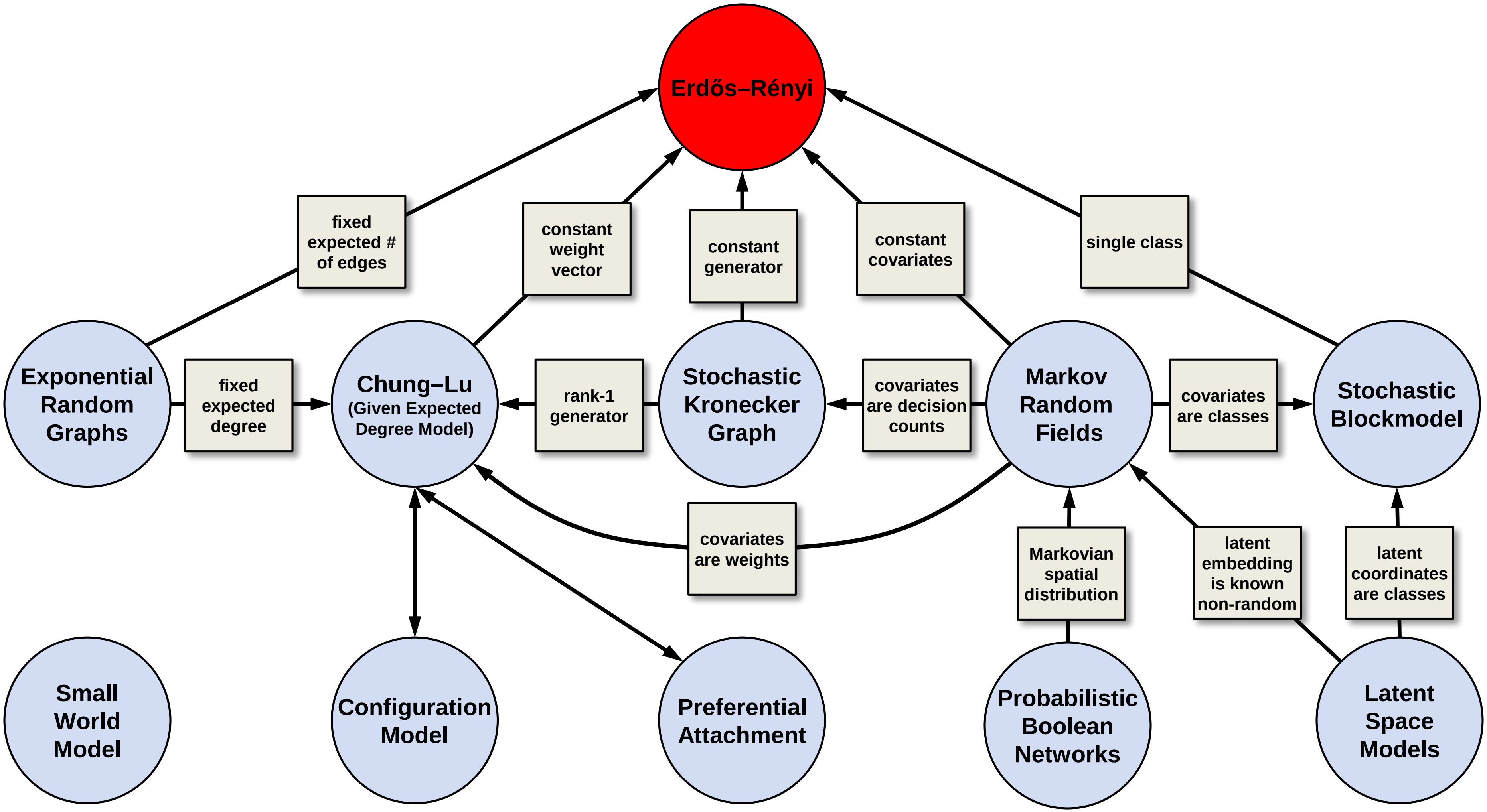
- Ex: 0^{th} order (independent) factorization: $\pi_i = \{x_i\}$ or $\eta_i = \{e_i\}$ (singletons)

$p(x_n) \cdots p(x_2)p(x_1) = \prod_{j=1}^{n} f_j(x_j)$, (Attributional factor graph)

$p(e_P) \cdots p(e_2)p(e_1) = \prod_{j=1}^{P} f_j(e_j)$, (Relational factor graph)

Ref: Loeliger 2007, Lauritzen 1996

Generative random graph models



Adapted from Nicholas Arcolano "Statistical models for random graphs" 2011.

Gilbert-Erdős-Rényi (ER) random graphs

- A special case of a Bernoulli graph
  - Every edge *e* has two states {0, 1}
  - $P(a_{ij} = 1) = \theta_{ij} = \theta$ (all edges equally likely)

- Introduced by R. Solomonoff and A. Rapoport (1951), E. N. Gilbert (1959)
- P. Erdős and A. Rényi(1959) model: *m* edges randomly and uniformly distributed among *n* nodes ($\theta = \binom{n}{m}^{-1}$ as n,m→ ∞)
- Summary statistics
  - Mean # of edges = $\binom{n}{2}\theta$
  - Mean degree = $(n-1)\theta$
  - Binomial degree distribution:

$$P(d_i = k) = \binom{n-1}{k}\theta^k(1-\theta)^{n-1-k}$$

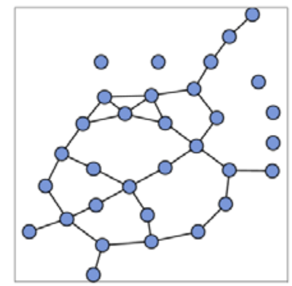
  - Degree converges to Poisson as $n \to \infty, \theta \to 0, n\theta = \lambda$

$$P(d_i = k) \to \frac{\lambda^k}{k!}\exp(-\lambda)$$

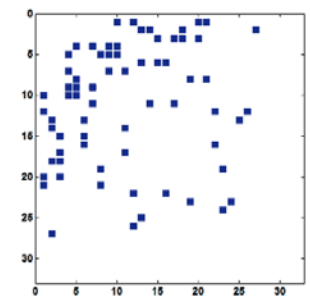
- ML estimator of θ is closed form
  - $\hat{\theta} = \frac{2m}{n(n-1)} \Rightarrow$ normal distributed as $p \to \infty$

$$p(e_{n,n-1}) \cdots p(e_{1,3})p(e_{1,2}) = \prod_{i>j} f_{ij}(e_{i,j})$$

$$f_{ij}(e_{ij}) = \theta^{e_{ij}}(1-\theta)^{1-e_{ij}}, \quad \theta \in [0,1]$$

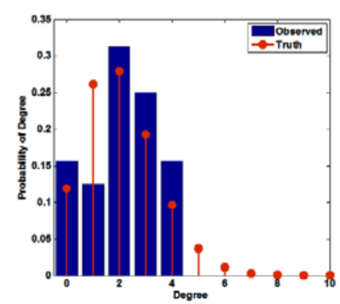


Example of Graph Generated from Model
(32 nodes, 34 edges)



Adjacency Matrix of Graph
(nnz = 68, sparsity = 0.0625)

$$\theta = 0.0645, n\theta = 2.06$$

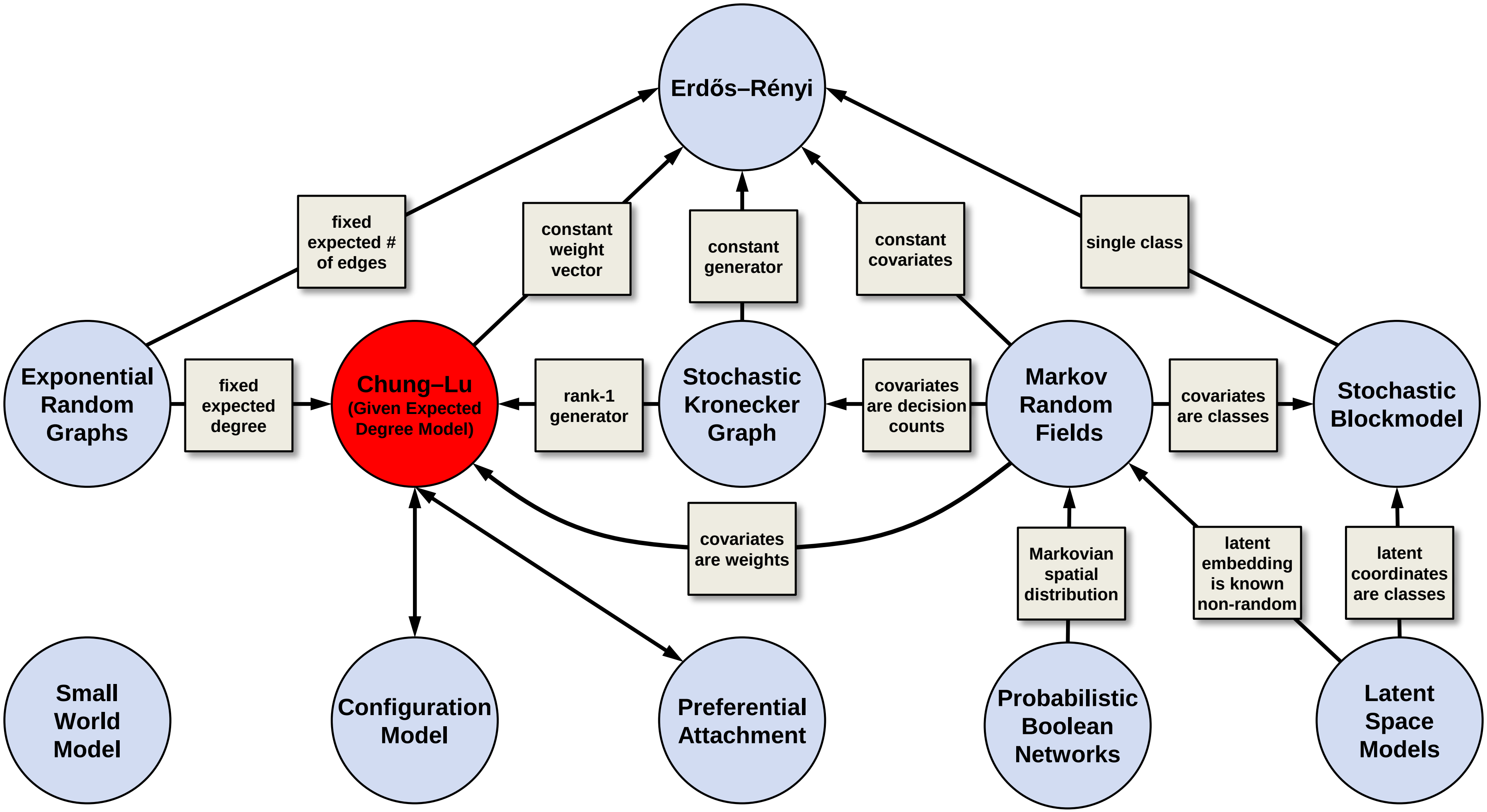


Observed and Theoretical Degree Distributions
(normalized; linear scale)

Arcolano 2011

Kolaczyk 2009

Generative random graph models



Adapted from Nicholas Arcolano “Statistical models for random graphs” 2011.

Chung-Lu random graphs

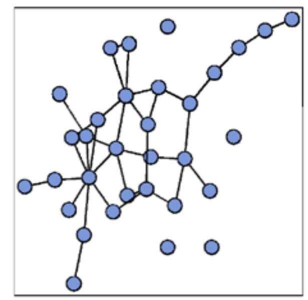
- [Chung-Lu 2002] edges are Bernoulli with
  - $P(a_{ij} = 1) = \theta_{ij} = \omega_i \omega_j$
  - $\boldsymbol{\omega} = [\omega_1, \dots, \omega_n]^T \in [0,1]^n$ is a weight vector
  - $E[\boldsymbol{A}] = \boldsymbol{\omega}\boldsymbol{\omega}^T \Rightarrow$ mean adjacencies $E[a_{ij}] = \omega_i \omega_j$
- Each node i has mean degree

$$d_{avg}(i) = E[d_i] = \omega_i ||\boldsymbol{\omega}||_1, \ \text{i} = 1, \dots, p$$

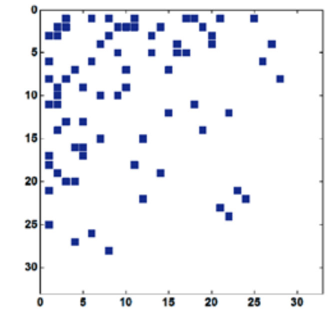
- Overcomes some Erdős-Rényi deficiencies
  - Probability of an edge varies over network
  - Degree distribution approximates power law
  - Induces small world properties
- Parameter estimation:
  - ML estimator not closed form
  - MoM estimators are often used instead
- Some SP applications of Chung-Lu
  - Anomaly detection in social networks [Miller 2013]
  - Modeling biological networks [Chung 2003]

$$p(e_{n,n-1}) \cdots p(e_{1,3}) p(e_{1,2}) = \prod_{i>j} f_{ij}(e_{i,j})$$

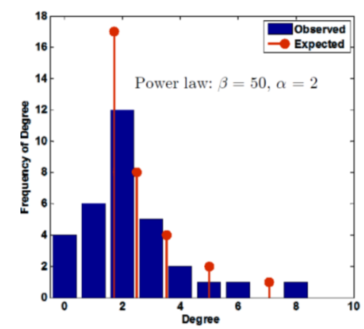
$$f_{ij}(e_{ij}) = \theta_{ij}{}^{e_{ij}} (1 - \theta_{ij})^{1 - e_{ij}}, \qquad \theta_{ij} = \omega_i \omega_j$$



Example of Graph Generated from Model
(32 nodes, 38 edges, 2 self-edges)



Adjacency Matrix of Graph
(nnz = 74, sparsity = 0.0723)

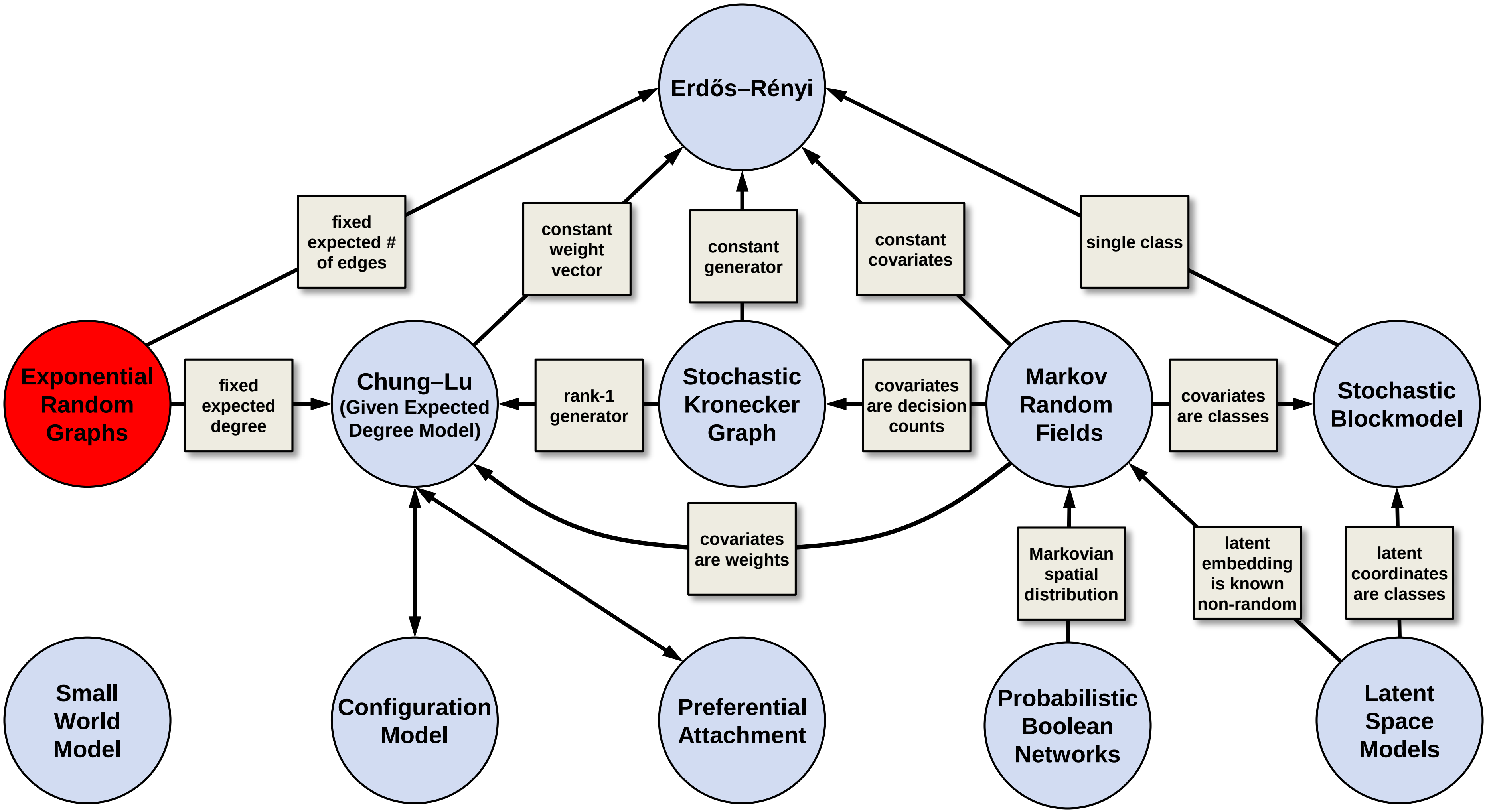


Observed and Expected Degree Distributions
(unnormalized; linear scale)

Arcolano 2011

Kolaczyk 2009

Generative random graph models



Adapted from Nicholas Arcolano "Statistical models for random graphs" 2011.

Exponential random graph model (ERGM)

- Erdős-Rényi and Chung-Lu models are both completely specified by their mean degrees $E[d_i] = (n-1)\theta$ and $E[d_i] = \omega_i||\boldsymbol{\omega}||_1$, resp.
- What if wanted a model that matched M specified moments?
- Moment constraints on model $P(G)$:

$$E[g_i(G)] = \sum_{G\in\Omega_n} g_i(G)\, P(G) = y_i, \qquad i = 1, \dots, M$$

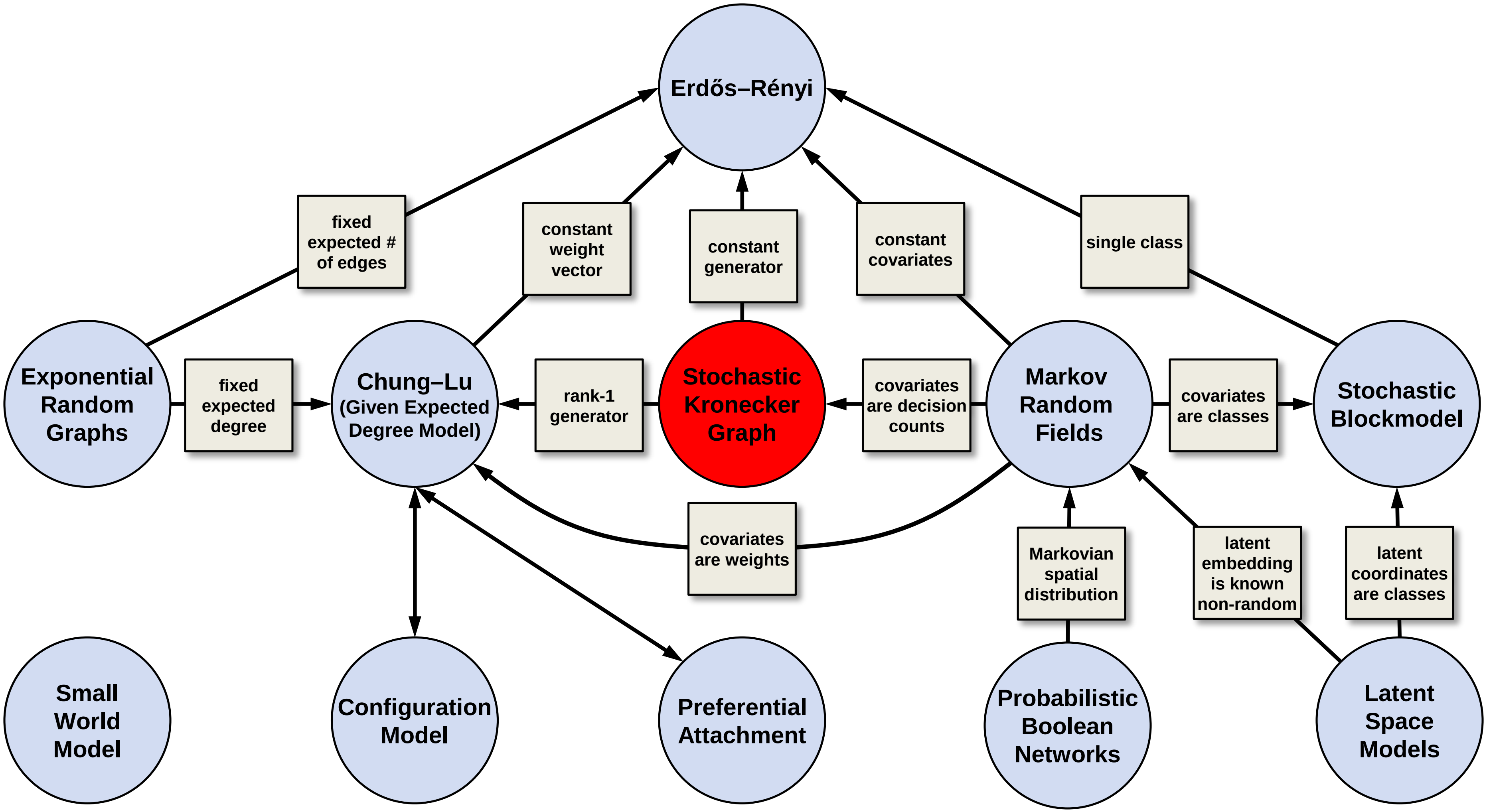
- Philosophy behind exponential random graphs: select solution $P^*(G)$ that maximizes entropy while satisfying constraints
- Maximum entropy solution has well known form [Kolaczyk 2009]

$$P(G) \propto \exp\left(\sum_{i=1}^{M} \beta_i(y_i) g_i(G)\right)$$

- For M=1, $g_1(G)$= number of edges obtain Erdős-Rényi model
- For M=p, $g_i(G)$=degree of *i*-th node obtain Chung-Lu model

Kolaczyk 2009

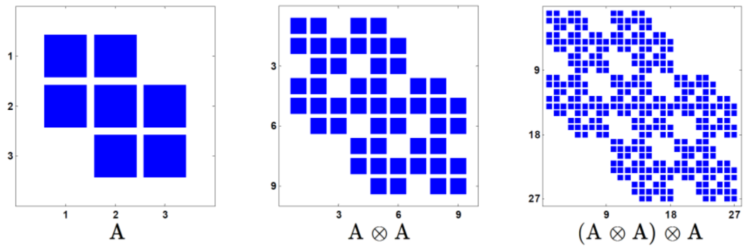
Generative random graph models



Adapted from Nicholas Arcolano "Statistical models for random graphs" 2011.

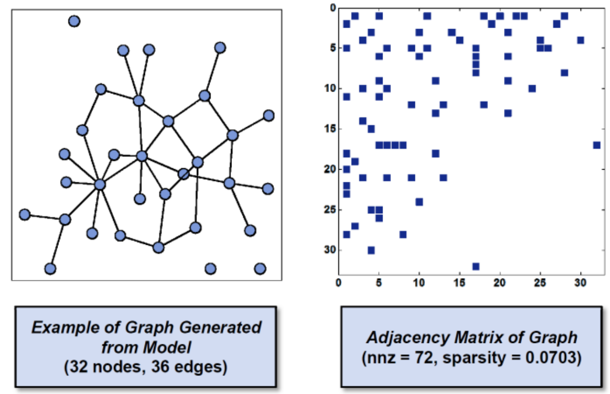
Stochastic Kronecker random graphs

- Proposed by [Leskovec 2005] as a way to better control degree distribution
- Edge probability matrix $\Theta = ((\theta_{ij}))$ generated recursively by Kronecker mult.

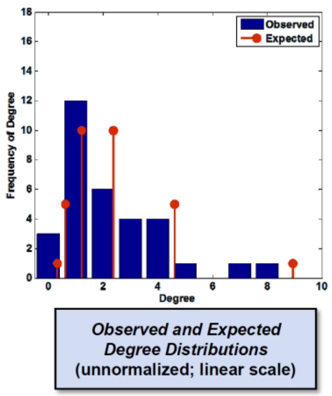


- Can so generate very large stochastic Kronecker graphs [Chakrabarti 2004]
- Can infer generator generator A from Θ using MCMC [Leskovec 2010]
- Global degree distribution is multinomial
- For large graphs diameter of graph is constant with high probability
- Good fit to real data [Leskovec 2010]
- Kronecker vs. Chung-Lu? [Pinar 2011]

Example

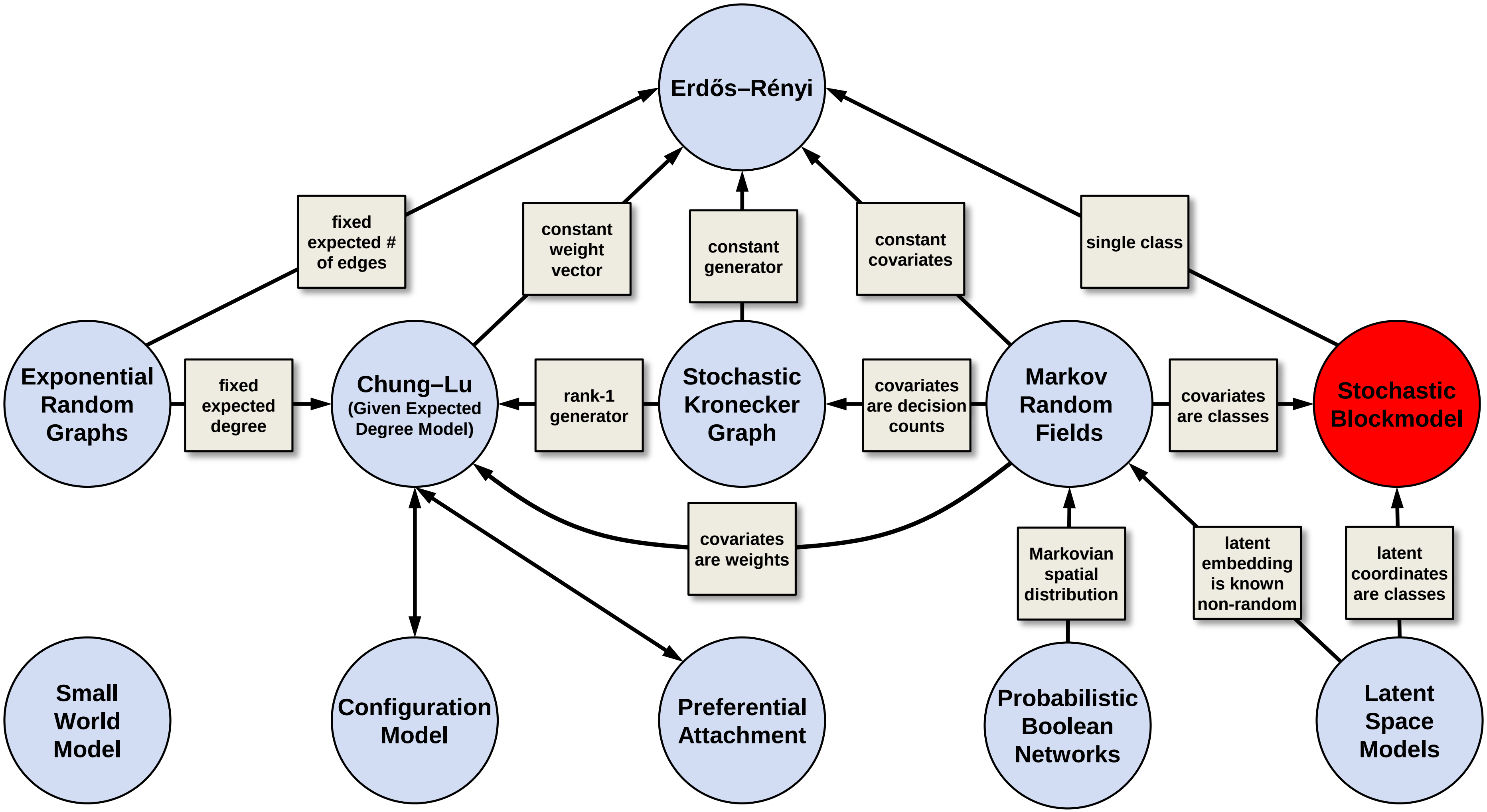


$n = 32, k = 5, \alpha = 0.95,$
$\beta = 0.60, \gamma = 0.20$



Arcolano 2011

Generative random graph models



Adapted from Nicholas Arcolano "Statistical models for random graphs" 2011.

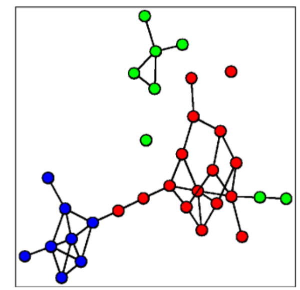
Stochastic block model (SBM)

- SBM is a multiclass extension of Erdős-Rényi [Wang 1987]
- A community detection and clustering method
- SBM is a LSM where latent variables $Z_i \in \{1, \dots, q\}$ are hidden class attributes of the nodes
- Divides adjacency matrix into blocks according to node classes and induces *stochastic equivalence* between nodes in the same class
- Probability model: define ρ_{kl}=probability of connection between classes *k* and *l*

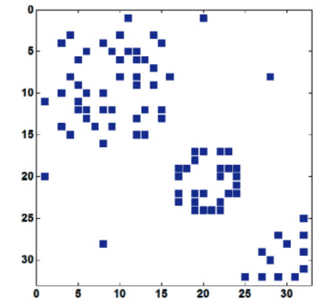
$$P(a_{ij} = 1 | Z_1, \dots, Z_n) = P(a_{ij} = 1 | Z_i, Z_j) = \rho_{Z_i, Z_j}$$

- Fitting model: EM, logistic lasso, MC , etc
  - A priori model: estimate all $\{\rho_{kl}\}$
  - A posteriori model: estimate $\{\rho_{kl}\}, \{Z_i\}$
- Applications include
  - Biological/socialnetworks [Ahmed 2009]
  - Geopolitical networks [Westveld 2011]
  - Dynamical social networks [Xu 2014],

Example:

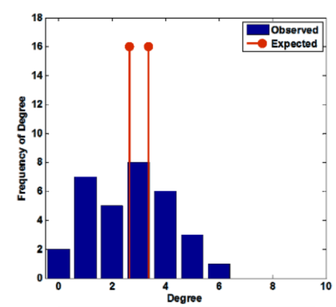


Example of Graph Generated from Model (32 nodes, 43 edges)



Adjacency Matrix of Graph (nnz = 86, sparsity = 0.0840)

$$n = 32, q = 3, n_1 = 16, n_2 = n_3 = 8$$
$$\rho_{11} = 0.2, \rho_{22} = 0.3, o.w.\ \rho_{ij} = 0.01$$

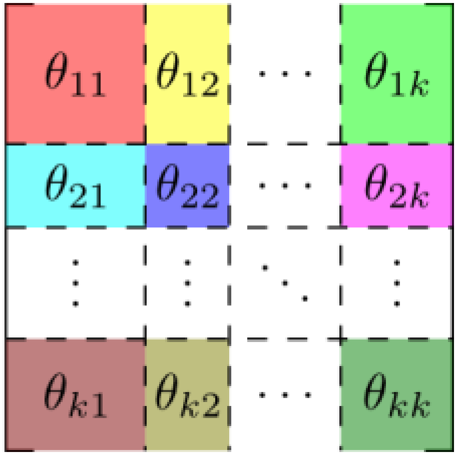


Observed and Expected Degree Distributions (unnormalized; linear scale)

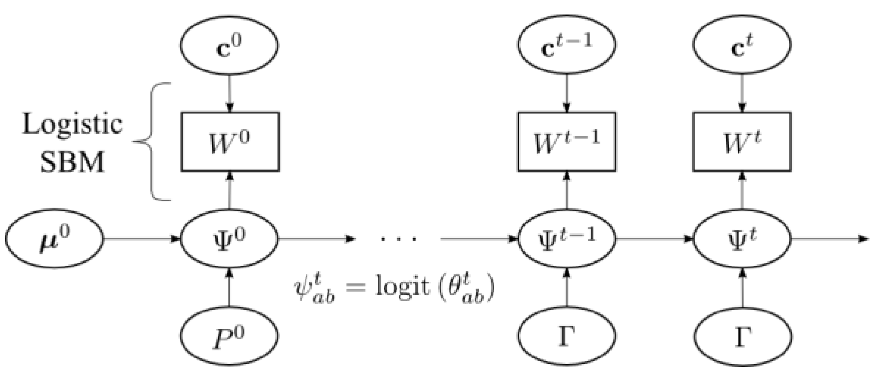
Arcolano 2011

Dynamic SBM [Xu, Kliger, H 2014]

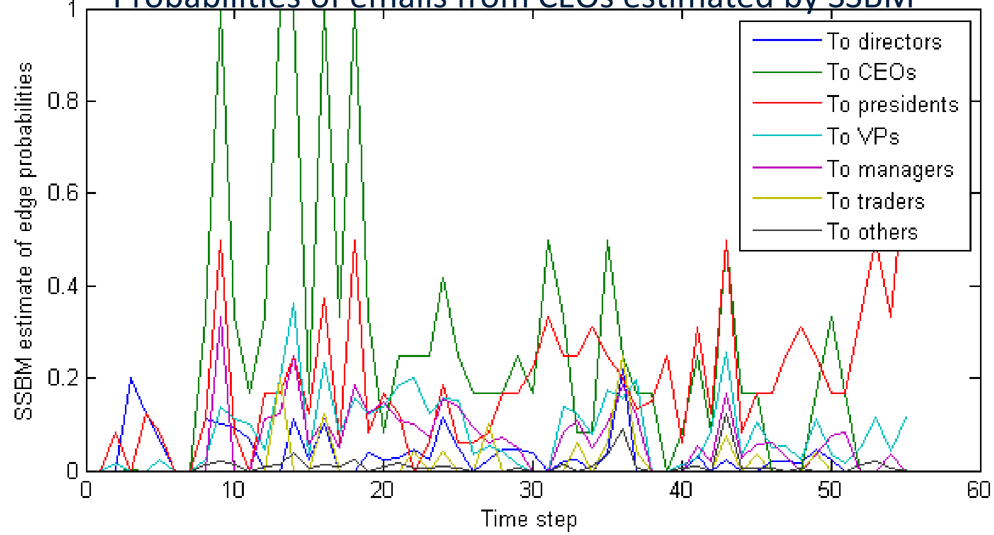
$\Theta^t \sim$



$$\Psi^t = \text{logit}(\Theta^t) = \log(\Theta^t) - \log(1 - \Theta^t)$$

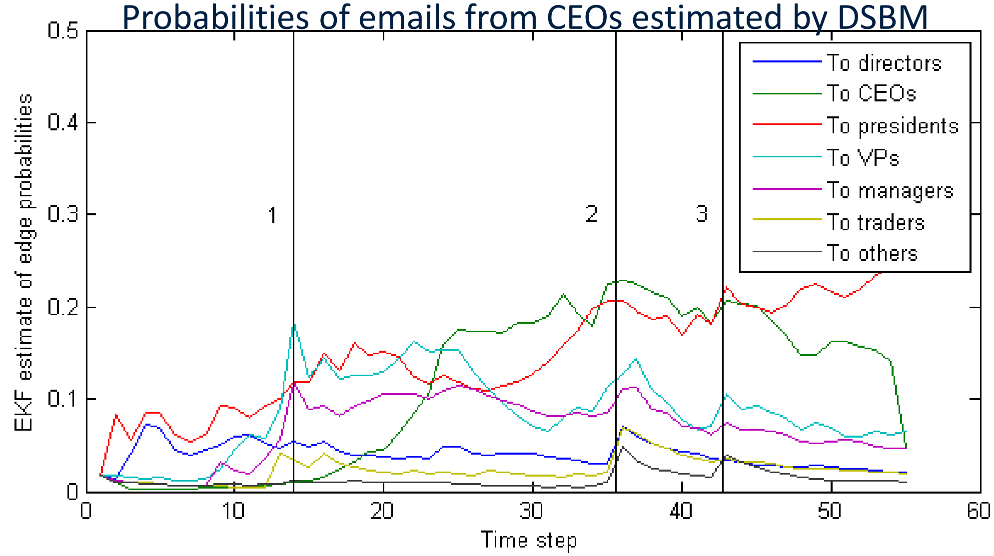


Probabilities of emails from CEOs estimated by SSBM

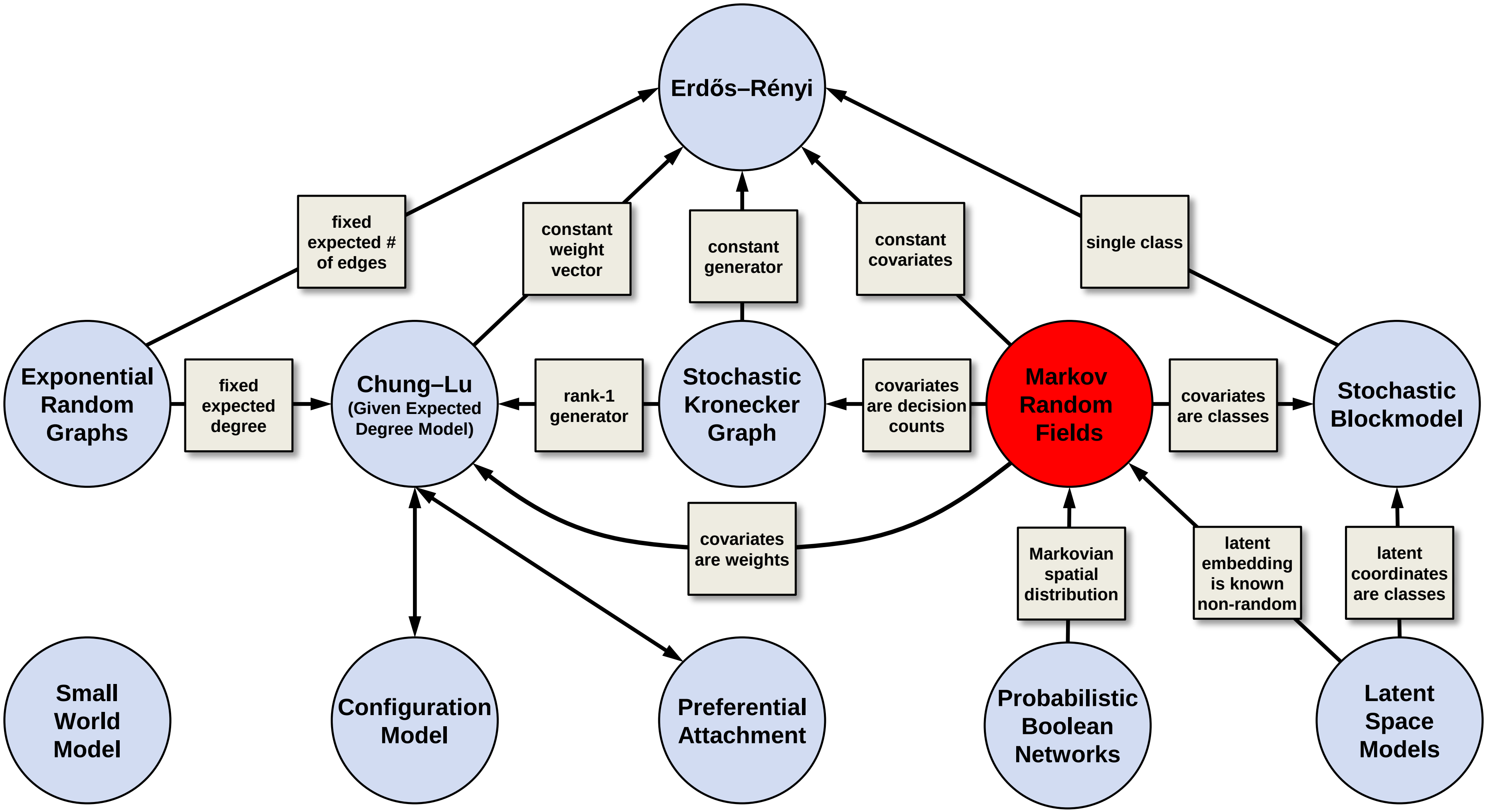


1. Enron issues Code of Ethics
2. Enron's stock closes below $60
3. CEO Skilling resigns

Probabilities of emails from CEOs estimated by DSBM

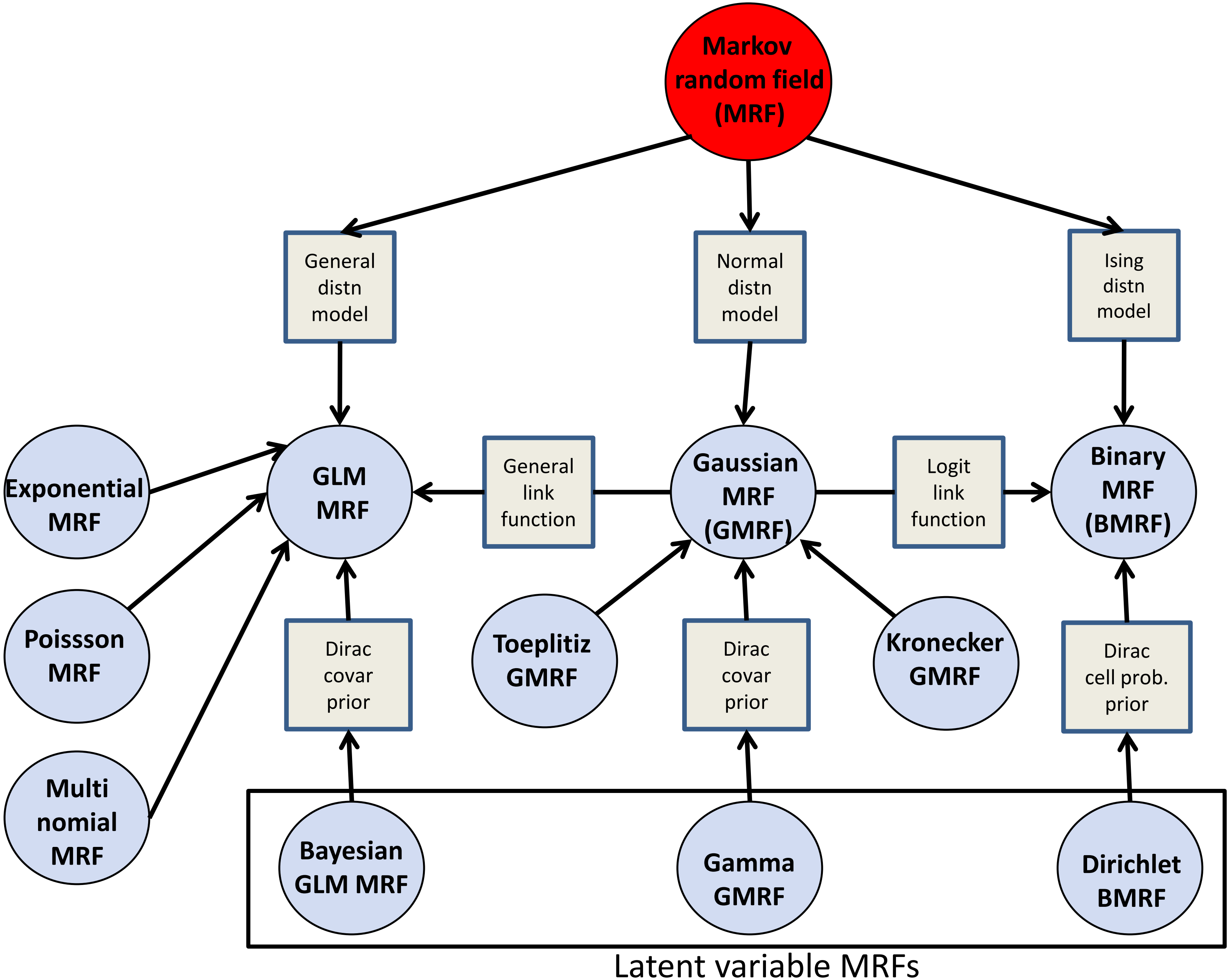


Generative random graph models



Adapted from Nicholas Arcolano "Statistical models for random graphs" 2011.

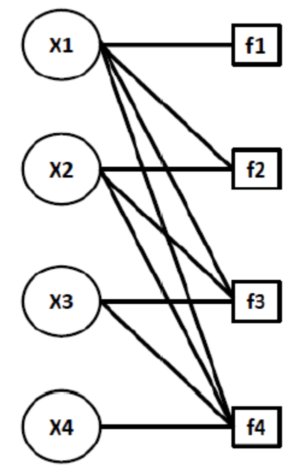
Subclasses of MRF random graph models



Go back to factor graphs: nodal examples

- Universal factorization ("saturated model"): $\pi_{i+1} = \{x_{i+1}, \pi_i\}$
  - $\pi_1 = \{x_1\}, \pi_2 = \{x_2, x_1\}, \dots, \pi_n = \{x_n, \dots, x_1\}$

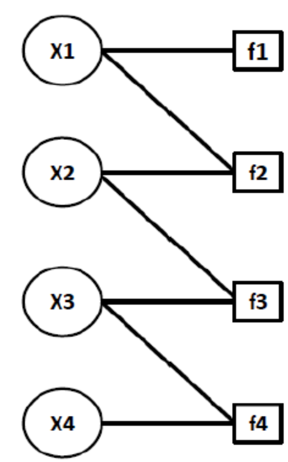
$$p(x_n|x_{n-1}, \dots, x_1) \cdots p(x_2|x_1)p(x_1) = \prod_{j=1}^{N} f_j(x_{\pi_j})$$



saturated model.

- 1st order (Markov) factorization: $\pi_{i+1} = \{x_{i+1}, x_i\}$
  - $\pi_1 = \{x_1\}, \pi_2 = \{x_2, x_1\}, \dots, \pi_p = \{x_n, x_{n-1}\}$

$$p(x_n|x_{n-1}) \cdots p(x_2|x_1)p(x_1) = \prod_{j=1}^{N} f_j(x_{\pi_j})$$

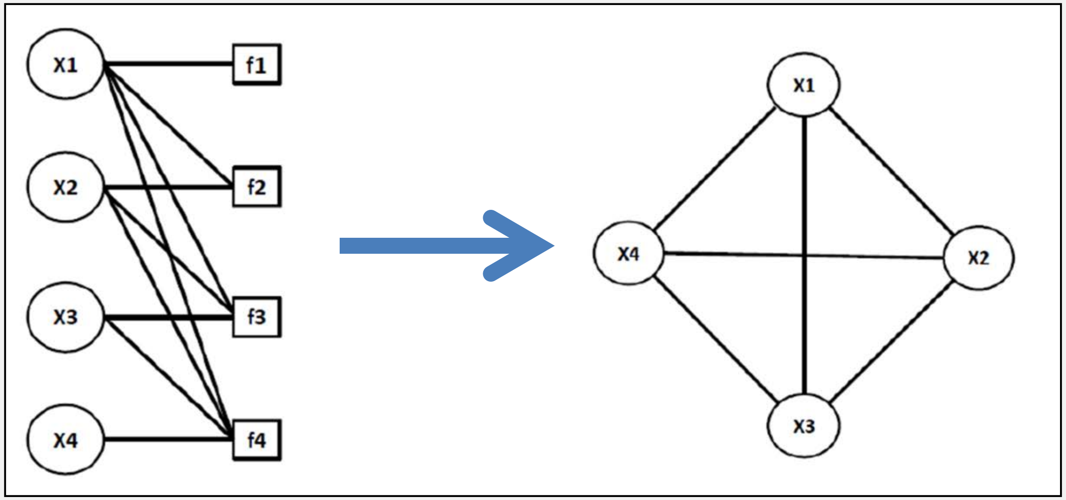


Markov model

Nodal factor graphs and Markov graphs

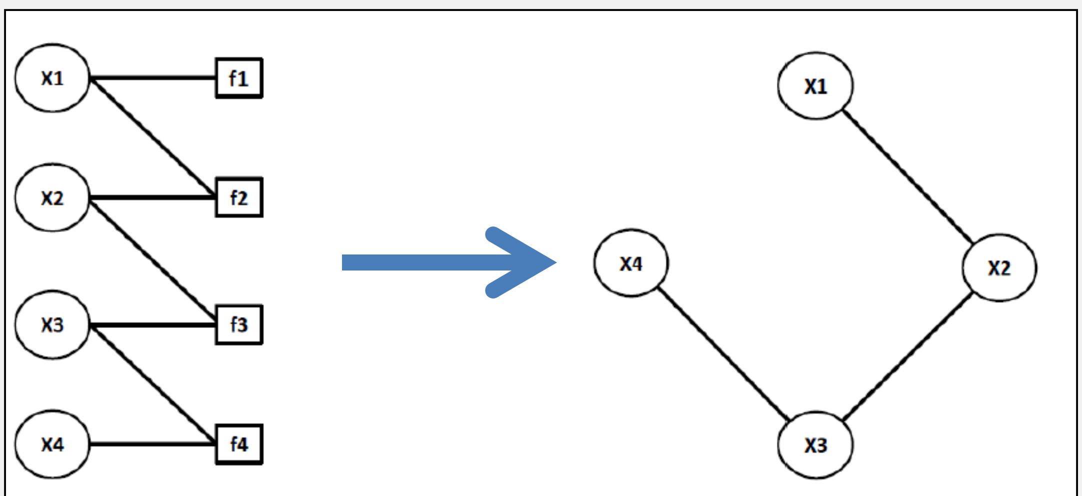
- A factor graph can be converted into a Markov graph by hiding the boxes f1

Saturated model



Complete graph

1st order Markov model



1st order Markov model

Factor graph

Markov graph

What is Markovian about a Markov graph?

- A Markov graph, also called a Markov random field (MRF), represents the conditional dependencies of the jpdf
- Let $C = \{\pi_j\}_{j=1}^{N}$ denote set of cliques of G
- G-compatible factorization of jpdf:

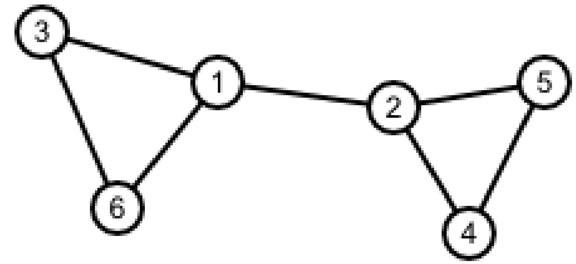
$$p(x_1, \dots, x_p) = \prod_{c \in C} f_c(x_c)$$

- Pairwise Markov property: for any pair a,b of non-adjacent nodes of G

$$p(x_a, x_b | x_{V \setminus ab}) = p(x_a | x_{V \setminus ab})\, p(x_b | x_{V \setminus ab})$$

- **Hammersley-Clifford theorem** [Speed 1979]: a positive jpdf satisfies the pairwise Markov property wrt G iff it has a G-compatible factorization.

Example: $G = (V, E)$



- N=3 cliques: $\{1,3,6\}, \{1,2\}, \{2,4,5\}$
- Pairwise Markov property

$$p(x_4, x_6 | x_{V \setminus 4,6}) = p(x_4 | x_{V \setminus 4,6}) p(x_6 | x_{V \setminus 4,6})$$

- G-compatible factorization

$$p(x_1, \dots, x_6) = p(x_3, x_6 | x_1) p(x_1, x_2) p(x_4, x_5 | x_2)$$

Ref: Lauritzen 1996

Markov Random Fields

- Several special cases of MRFs studied
  - Gaussian Markov random fields
  - Binary Markov random fields
  - Multinomial Markov random fields
  - Poisson Markov random fields
  - Generalized linear model MRFs

- Other names for MRFs:
  - Gibbs field (when probability > 0)
  - (Undirected) probabilistic graphical model

- Latent MRF: jpdf is described as integral over latent variables (hidden states) of a conditional jpdf given those variables.

- General model often difficult to apply directly
  - No general closed form representation exists for marginal distributions
  - Inference methods: MC, VB, EM, Gibbs, BP
  - Much work on tractable special cases [Koller 2009, Yang 2012]

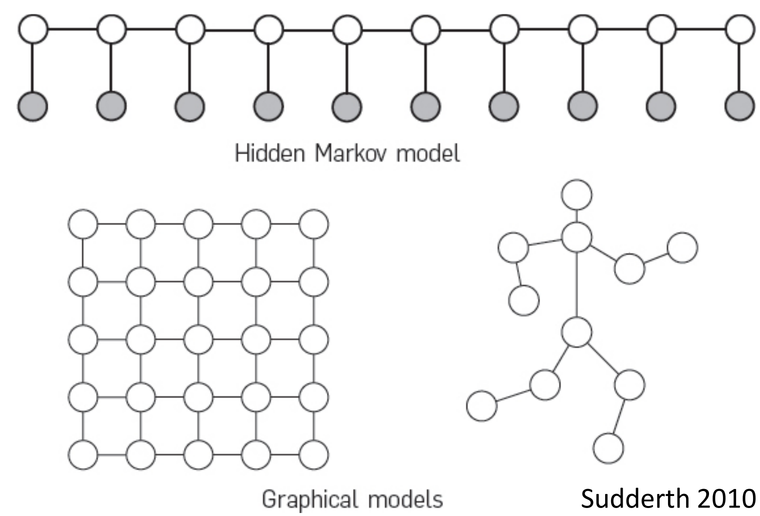
For any positive G-compatible jpdf

$$p(x_1, \ldots, x_p) = \exp\left(\sum_{c \in C} \theta_c U(x_c) - D(\theta)\right)$$

$U(x_c)$ are clique-wise sufficient statistics

Special case: pairwise interaction MRF

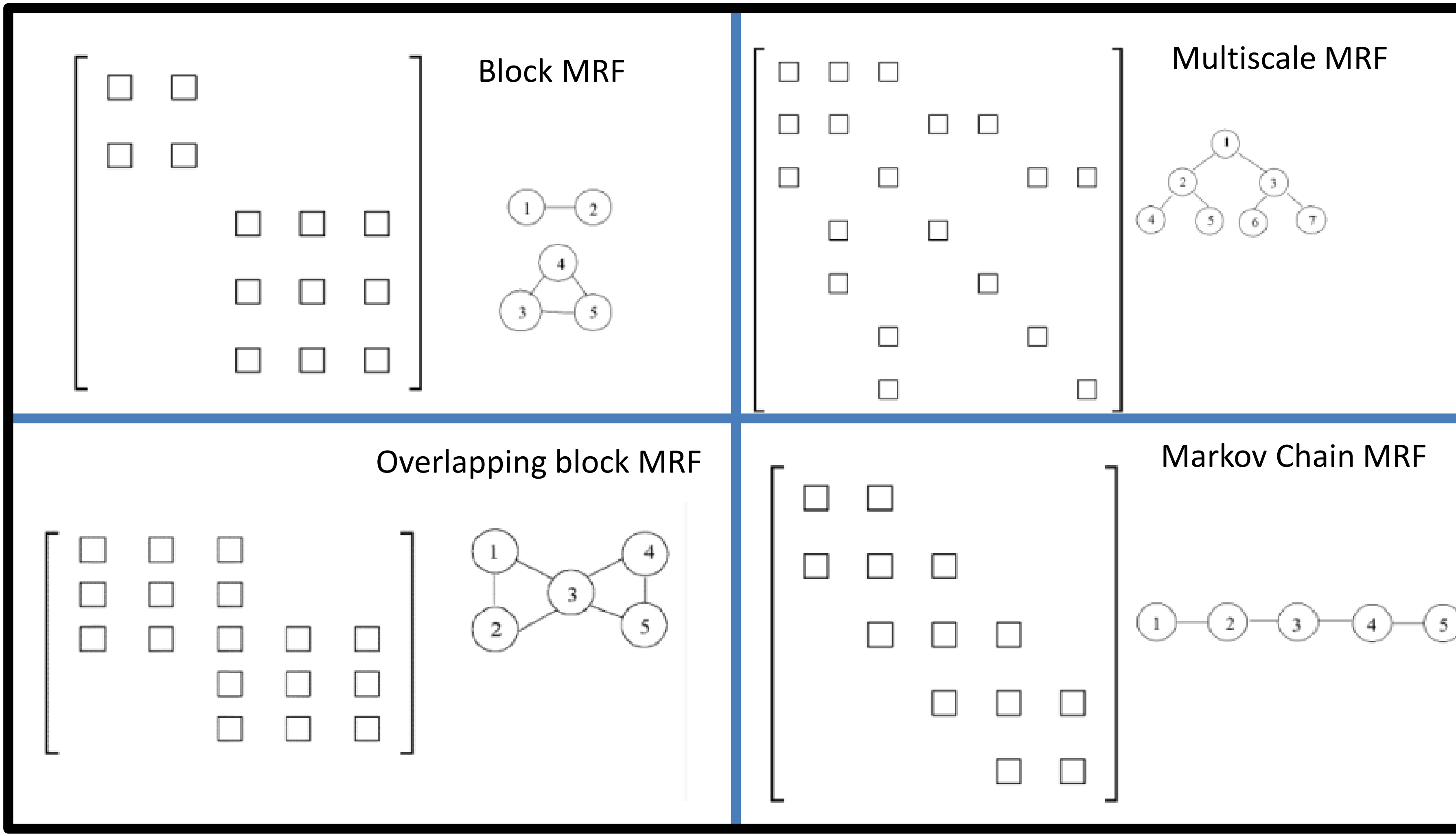
$$p(x_1, \ldots, x_p) \propto \exp\left(\sum_{i,j \in E} \theta_{ij} U(x_i, x_j)\right)$$



Sudderth 2010

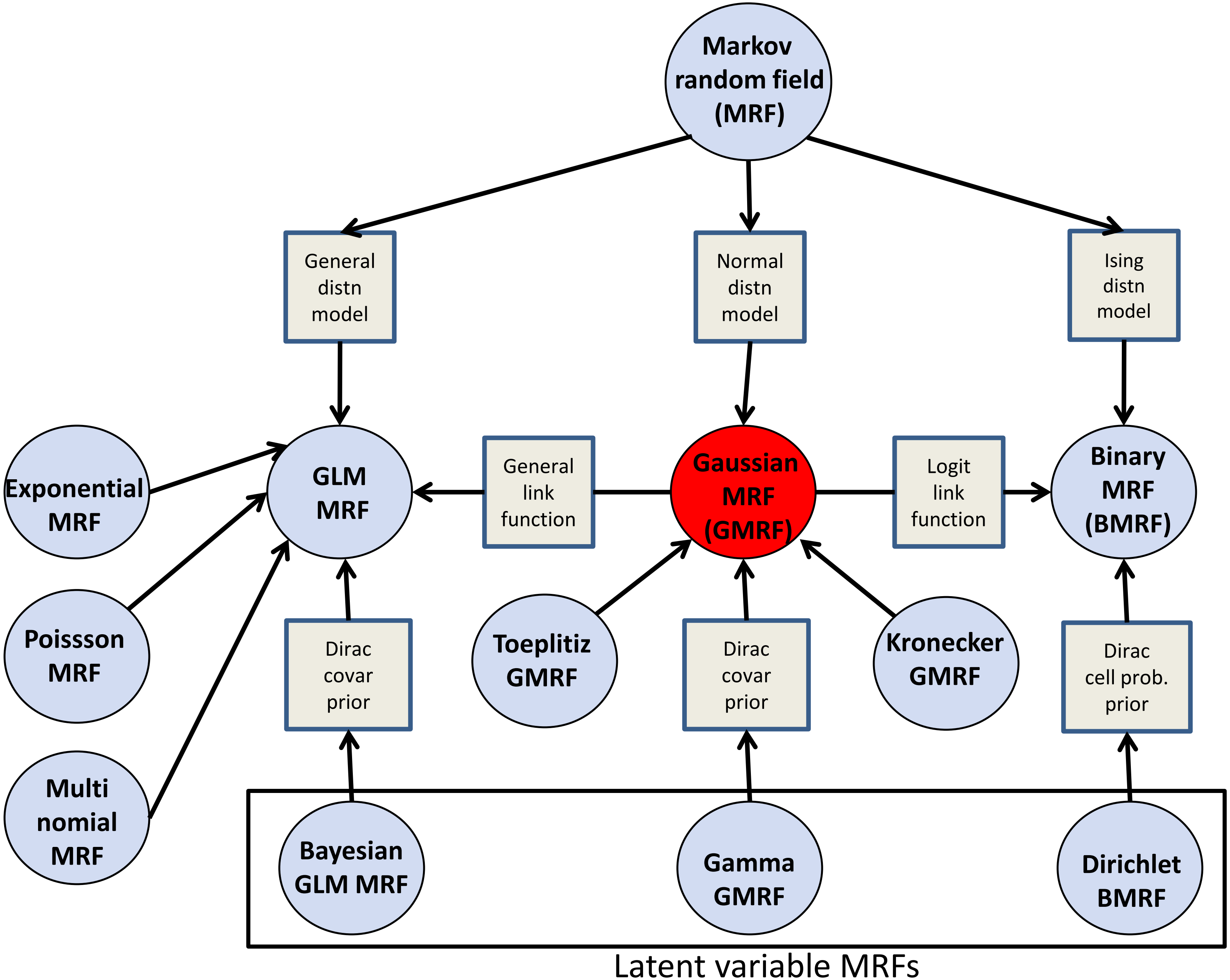
Lauritzen (1976). *Graphical Models*
Rue, Håvard; Held, Leonhard (2005). *Gaussian Markov random fields: theory and applications*.

Selected MRFs and their adjacency matrices



Source: Wiesel *et al* 2010

Subclasses of MRF random graph models



Gauss Markov Random Fields

- A MRF where the jpdf is Gaussian
  – All conditional and marginal distributions are Gaussian
- Edges in the graph are specified by non-zero entries in the inverse covariance (precision) matrix $\boldsymbol{K} = \boldsymbol{\Sigma}^{-1}$
- Estimation of GGM: N>n i.i.d. samples of $[x_1, \ldots, x_n]$
  – ML estimator of $\boldsymbol{K}$ is $\boldsymbol{S}_N^{-1}$, sample covariance
- Estimation of GGM: n<p i.i.d. samples of $[x_1, \ldots, x_n]$
  – Lasso nodewise regression [Meinshausen 2006]
  – Glasso [Friedman 2007] sparse MLE of $\boldsymbol{K}$
  – Thresholded Moore-Penrose Z-score [Hero 2011, 2012]
- Structure on $\boldsymbol{K}$ is often imposed to handle n<p
  – Toeplitz [Bach 2004] stationary
  – Sparse+Kronecker [Tsiligkaridis 2013] spatio-temporal
  – Sparse+Toeplitz+Kronecker [Greenewald, H 2014]
- Latent variable extension - use conjugate prior for $\boldsymbol{K}$: inverse Gamma distribution [Rajaratnam 2008]
- Applications:
  – image segmentation [Willsky 2002]
  – Computer vision [Li 2007]
  – Biological networks [Friedman 2004]
  – Spatio-temporal [Greenewald 2013, Firouzi 2014]

- Joint pdf: $p(\boldsymbol{x}) = \dfrac{\exp\{-\boldsymbol{x}^T \boldsymbol{K}\, \boldsymbol{x}/2\}}{\kappa(\boldsymbol{K})}$

$$-\boldsymbol{x}^T \boldsymbol{K}\, \boldsymbol{x} = \sum_{i \in \mathrm{V}} \omega_{ii} x_i^2 + \sum_{i,j \in E} \omega_{ij} x_i x_j$$

- ω_{ij} related to partial correlation

$$\rho_{ij\,|\,V\backslash ij} = -\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}$$

- ω_{ij} and conditional independence

$$\rho_{ij\,|\,V\backslash ij} = 0 \Leftrightarrow x_i, x_j \perp x_{V\backslash ij}$$

- ω_{ij} and node-wise regression

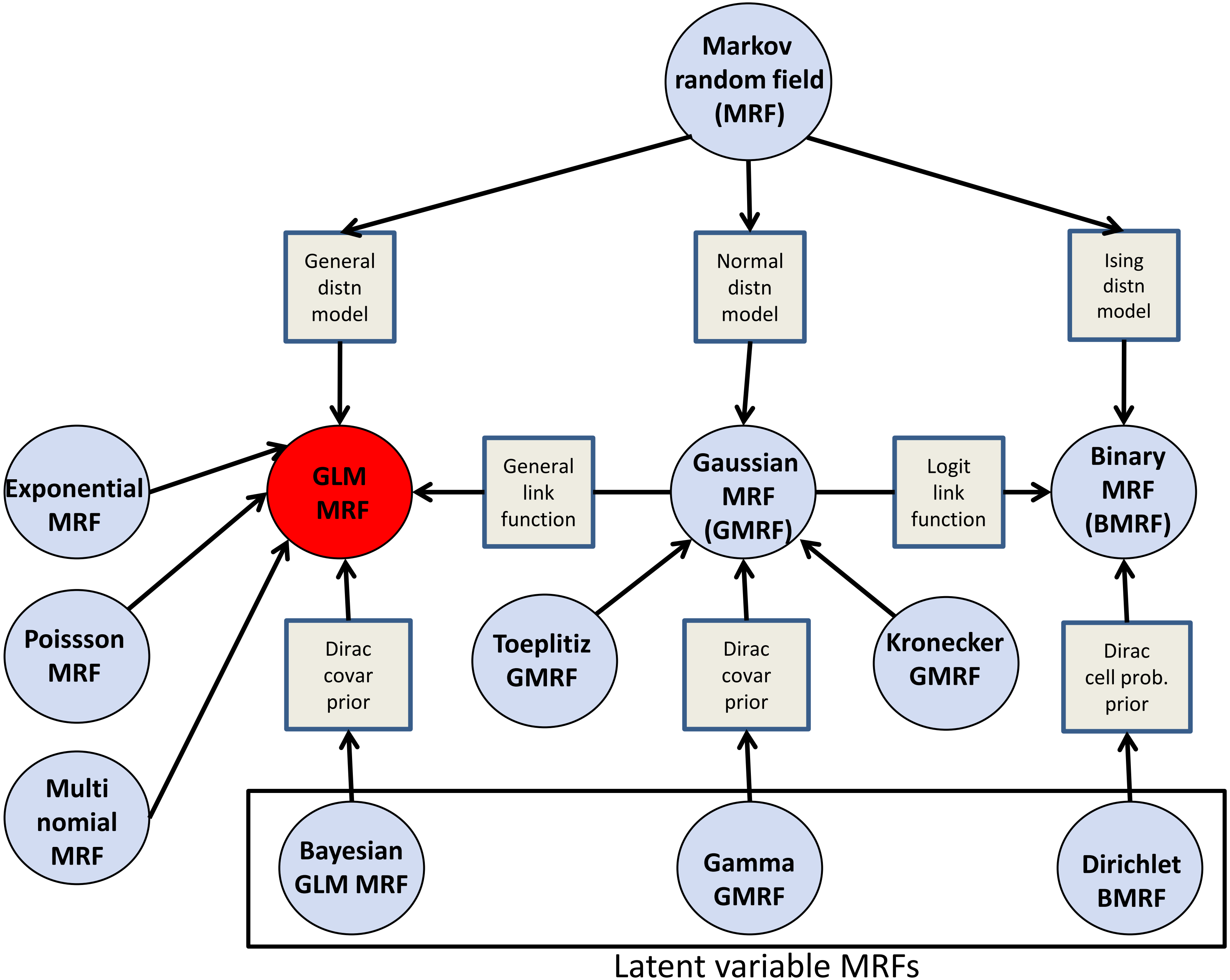
$$x_i = \sum_{j \neq i} \beta_{ij} x_j + \epsilon_i \qquad \beta_{ij} = \frac{\omega_{ij}}{\omega_{ii}}$$

⇒ Suggests estimating MRF using lasso

$$amin\_\boldsymbol{\beta}_{i*} \left\{ \sum_j (x_i - \boldsymbol{\beta}_{i*}^T \boldsymbol{x}_{-i})^2 + \lambda ||\boldsymbol{\beta}_{i*}||_1 \right\}$$

[Meinshausen&Buhlmann 2006]

Subclasses of MRF random graph models



Generalized linear models for MRFs

- In principle iterative algorithms, e.g., MC, VB, message passing, can be applied to infer any MRF but are generally slow in high dimensions
- Some other approaches to non-Gaussian MRFs
  - Pre-transformations to Gaussian: $\log(x), \sqrt{x}$
  - Copula transformations on observations: [Liu 2009]
- Popular alternative: turn MRF inference into a prediction problem and use generalized linear model (GLM) to construct predictor [Nelder 1972]
- GLM principle: a certain transformed linear predictor is accurate
- GLM applied to MRFs of following types
  - Multinomial, Poisson, Exp [Yang 2013]

Elements of GLMs

Given response/predictor variables Y and X

1. $P(Y|\boldsymbol{X},\theta)$ conditional dsn of a response variable Y from exponential family with mean
$$\mu = E[Y|\boldsymbol{X},\theta]$$
2. A linear predictor
$$\eta = \boldsymbol{\beta}^T\boldsymbol{X}$$
3. A link function g such that
$$g(\mu) = \eta$$

Combining 2 and 3, GLM for (Y, X) is

$$g(E[Y|\boldsymbol{X},\theta]) = \boldsymbol{\beta}^T\boldsymbol{X} + \epsilon$$

Examples:

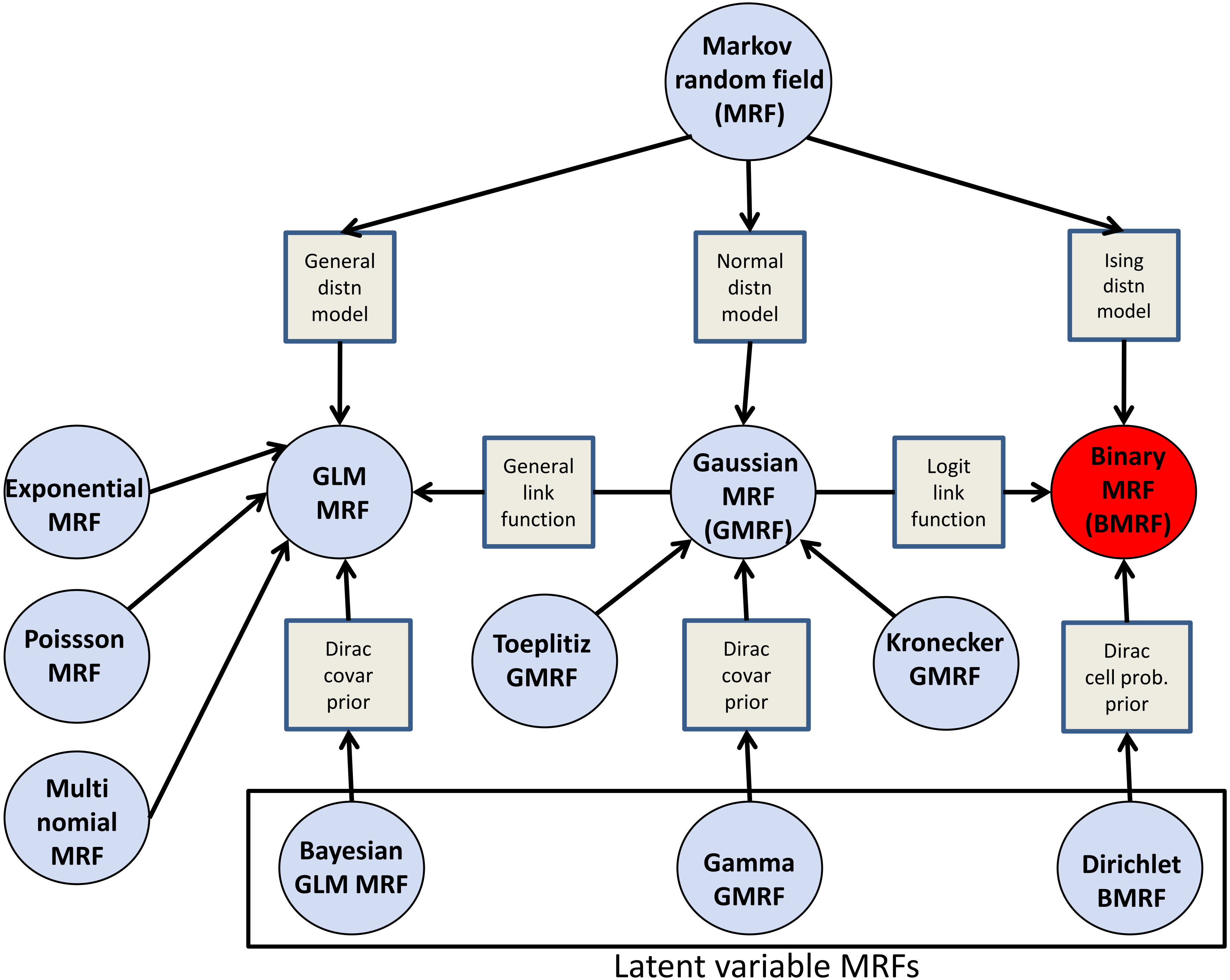
*Gaussian Y, X: $\mu = \theta,\ \ g(\mu) = \mu$

*Bernoulli Y, X: $\mu = \theta,\ \ g(\mu) = \log \mu/(1-\mu)$

*Poisson Y, X: $\mu = \theta,\ \ g(\mu) = \log \mu$

*Exponential Y, X: $\mu = \frac{1}{\theta},\ \ g(\mu) = -\mu^{-1}$

Subclasses of MRF random graph models

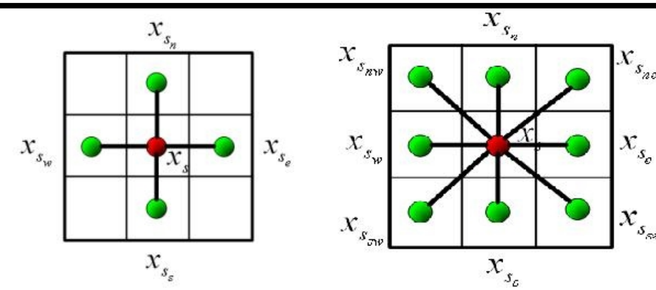


Binary Markov Random Fields

- A MRF where node attributes x_i are binary
- Introduced by Ising [Ising 1925]
  - Originally intended to model spin coupling in ferromagnetic materials
  - Basic neighborhood was local (1NN) in 1D
  - Generalizes to q-ary state Potts Model [Potts 1952]
- General form of jpdf

$$p(x_1, \dots, x_n) = \exp\left(\sum_{c \in C} \gamma_c U(x_c) - D(\gamma)\right)$$

- Inference of cliques C and parameter γ
  - MCMC and Gibbs sampling [Newman 1999]
  - Message passing [Wainright 2008]
  - L1 penalization (Glasso) [Ravikumar 2006]
  - Nodewise regression [Ravikumar 2010]
- Hierarchical Bayesian extensions
  - Topic models [Blei&Ng&Jordan 2003]
  - Dynamic Bayesian nets [Koller&Freidman 2009]
- SP examples
  - Image restoration [Besag 1991]
  - Image segmentation [Wainright 2008]



Pairwise Ising model joint pdf: $x_i \in \{0,1\}$

$$p(x_1, \dots, x_n) = \exp\left(\sum_{i,j=1} \gamma_{ij} x_i x_j - D(\gamma)\right)$$

γ_{ij} is related to conditional independence

$$\gamma_{ij} = \gamma_{ji} = 0 \Leftrightarrow x_i, x_j \perp x_{V\backslash ij}$$

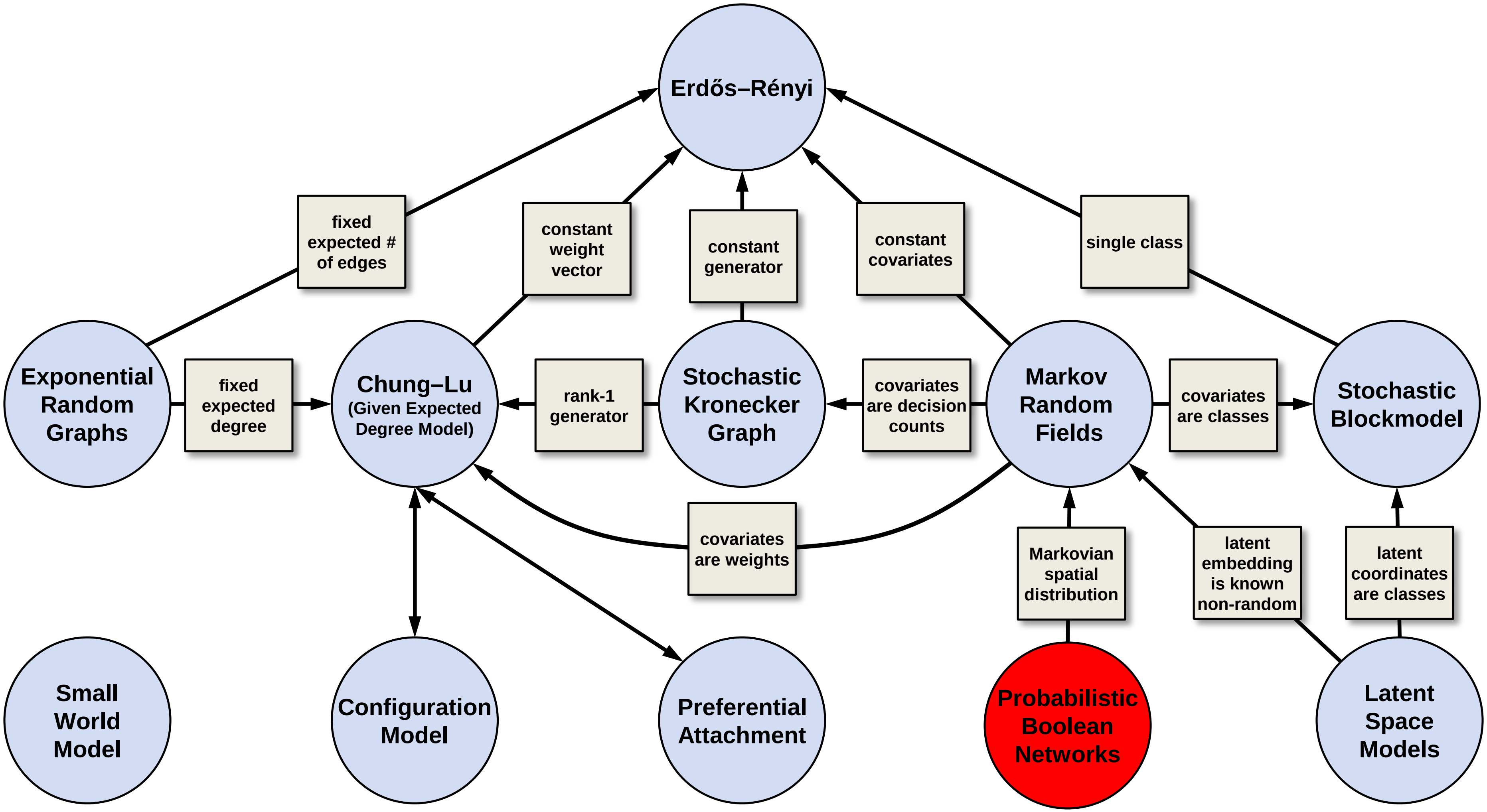
γ_{ij} is related to nodewise regression

$$\text{logit}(P(x_i = 1 | x_{V\backslash \text{i}})) = 2\sum_{j \neq i} \gamma_{ij} x_{ij} + \epsilon_i$$

Logit function : $\text{logit}(p) = \log(\frac{p}{1-p})$

$\Rightarrow$ logistic lasso for estimating MRF

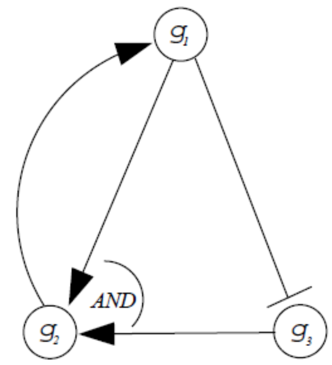
Attributional random graph models



Adapted from Nicholas Arcolano "Statistical models for random graphs" 2011.

Probabilistic boolean networks (PBN)

- Algebraic model introduced 45 years ago [Kauffman 1969] for rule-based dynamical systems described by Boolean state transition functions.
- Probabilistic Bayes nets (PBN) [Shmulevich 2002] to model uncertainty in data and in functional rules – posterior distribution over Boolean functions.
- Algebraic properties can be studied using Petri Nets [Steggles 2007, Karleback 2008]
- PBN = binary MRF [Murphy 1999] when synchronous, spatially Markov and acyclic.
- Inference: Markov Chains and MC
  - Petri net methods: detecting active pathways, reachability, state cycles, fixed points
  - Issues: scalability (2^p states), quantization
- Used extensively to model interacting binary systems in social science, statistical mechanics, network biology, automatic control, and signal processing communities.

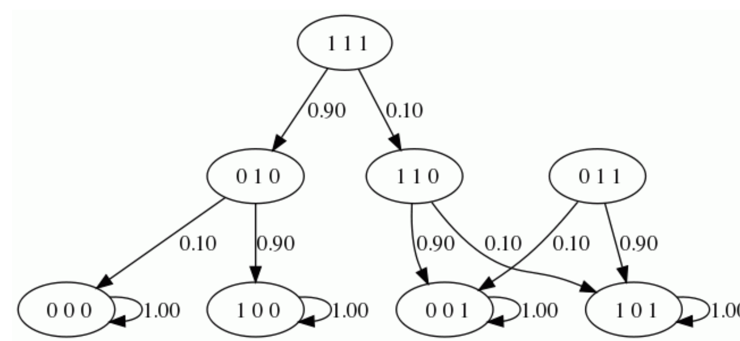


(a) Boolean network

Current			Next		
g_1	g_2	g_3	g_1'	g_2'	g_3'
0	0	0	0	0	1
0	0	1	0	0	1
0	1	0	1	0	1
0	1	1	1	0	1
1	0	0	0	0	0
1	0	1	0	1	0
1	1	0	1	0	0
1	1	1	1	1	0

(b) Truth table

Source: Steggles 2007



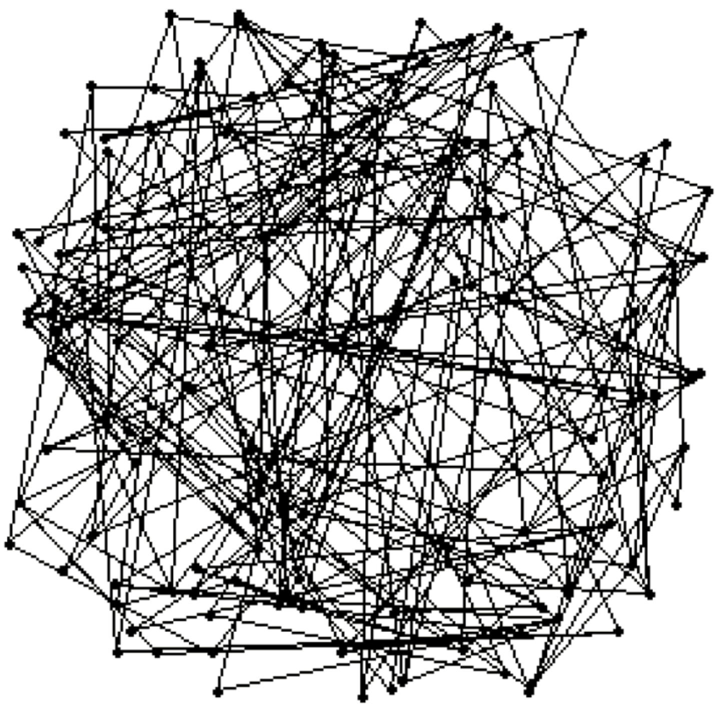
Source: http://adam.plantsimlab.org/userGuide.pl

General reference:
De Jong, H. (2002). Modeling and simulation of genetic regulatory systems: a literature review. *Journal of computational biology*, *9*(1), 67-103

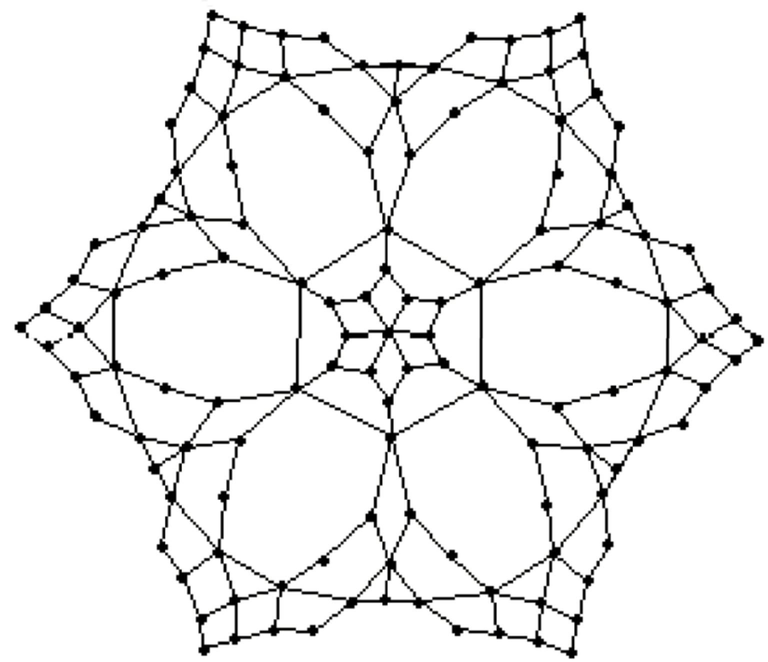
Wrapup

- "Signal processing for graphs:" modeling data using graphs
- Graph modeling is a rich area of practice and research
  - Summary statistics are useful for mining graphs and simple inference
  - Generative models are useful for simulation and complex inference
- There is an extensive toolbox of SP and statistical models
- Lots of emerging challenges for signal processing
  - Models that incorporate temporal dynamics and real-time updating
  - Models that include prior constraints reflecting patterns of adjacency
  - Models that combine different data types (relational, attributional)
  - Graph inference algorithms that are scalable to high dimensions
- Randomness can be in the eye of the beholder: layout is important!

Randomness can be in the eye of the beholder: layout is important!



Unweighted graph with nodes positioned randomly



Same graph with minimum energy layout (Davis 2009)

Source: Tim Davis' webpage. http://www.cise.ufl.edu/research/sparse/matrices/

Some software packages

- Bioinformatics toolbox. Matlab. http://www.mathworks.com/products/bioinfo/: graph functions and graph visualization.
- bnt (Bayes Net). Matlab – Murphy group, Google https://code.google.com/p/bnt/: static and dynamic Bayes nets, graph layout tools
- Statnet. Open source R package. http://csde.washington.edu/statnet/. ERGM models.
- HUGE. Open source R – Lafferty group, U. Chicago. Graph inference methods for Gaussian and related MRFs. On CRAN.
- BoolNet. Open source R. Package for generation, reconstruction and analysis of PBNs.
- Cytoscape. Open source with API. http://www.cytoscape.org/: data integration, analysis, and visualization
- ADAM (Analysis of Dynamic Algebraic Models) http://adam.plantsimlab.org – Laubenbacher Research Group at Virginia Tech: computational algebra for PBNs.

Selected References (1)

- [Ahmed 2009] A. Ahmed and E. P. Xing (2009). Recovering time-varying networks of dependencies in social and biological studies. *PNAS*, 106(29):11878–11883,
- [Bach 2004] Bach, F. R., & Jordan, M. I. (2004). Learning graphical models for stationary time series. *Signal Processing, IEEE Transactions on*, *52*(8), 2189-2199.
- [Besag 1991] Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, *43*(1), 1-20
- [Blei 2003] D. M. Blei, A. Y. Ng, and M. I. Jordan, (2003). Latent Dirichlet allocation,"*Journal of Machine Learning Research*, vol. 3, pp. 993–1022,.
- [Chen, H 2013] P.-Y. Chen and A.O. Hero, "Local Fiedler Vector Centrality for Detection of Deep and Overlapping Communities in Networks," *Proc. of IEEE Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, May 2014.
- [Chen, etal 2014] H. Chen, N. Comment, J. Chen, S. Ronquist, T Ried, A. Hero, I Rajapakse, "Chromosome Conformation of Human Fibroblasts Grown in 3-Dimensional Spheroids," submitted 2014.
- Chung, F. R. K. Spectral Graph Theory. Providence, RI: Amer. Math. Soc., 1997.
- [Coates, H, Nowak, Yu 2002] M Coates, AO Hero, R Nowak, B Yu, Internet Tomography, IEEE Signal Processing Magazine, Vol. 19, No. 3, pp. 47-65, May 2002.
- [Chung 2002] Chung, F., & Lu, L. (2002). The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, *99*(25), 15879-15882.
- [Chung 2003] Chung, F., Lu, L., Dewey, T. G., & Galas, D. J. (2003). Duplication models for biological networks. *Journal of computational biology*, *10*(5), 677-687.
- [Cross 1983] Cross, G. R., & Jain, A. K. (1983). Markov random field texture models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (1), 25-39.
- [De Jong 2002] De Jong, H. (2002). Modeling and simulation of genetic regulatory systems: a literature review. *Journal of computational biology*, *9*(1), 67-103.
- [Dimakis 2010] Dimakis, A. G., Kar, S., Moura, J. M., Rabbat, M. G., & Scaglione, A. (2010). Gossip algorithms for distributed signal processing. *Proceedings of the IEEE*, *98*(11), 1847-1864.
- [Firouzi 2014] Firouzi, H., Wei, D., & Hero III, A. O. (2014). Spectral Correlation Hub Screening of Multivariate Time Series.Excursions in Harmonic Analysis., 2014 Also available as *arXiv preprint arXiv:1403.3371*.
- M. Fiedler. Algebraic connectivity of Graphs, Czechoslovak Mathematical Journal: 23 (98), 1973.
- [Friedman 2004] Friedman, N. (2004). Inferring cellular networks using probabilistic graphical models. *Science*, *303*(5659), 799-805.
- [Friedman 2007] Friedman, J., Hastie, T. (2007).Sparse inverse covariance estimation with the graphical lasso, Biostatistics, 1-10.

Selected references (2)

- [Hero 2011] Hero, A., & Rajaratnam, B. (2011). Large-scale correlation screening. *Journal of the American Statistical Association*, *106*(496), 1540-1552.
- [Hero 2012] Hero, A., & Rajaratnam, B. (2012). Hub discovery in partial correlation graphs. *Information Theory, IEEE Transactions on*, *58*(9), 6064-6078.
- [Hero, Ma, Michel, Gorman 2002]. Applications of entropic graphs. *IEEE Signal Processing Magazine*, 19 (5), 85-95
- [Jordan 2004] Jordan, M. I. (2004). Graphical models. *Statistical Science*, *19*(1), 140-155.
- [Kang 2011] Kang, U., Chau, D. H., & Faloutsos, C. (2011, April). Mining large graphs: Algorithms, inference, and discoveries. In *Data Engineering (ICDE), 2011 IEEE 27th International Conference on* (pp. 243-254).
- [Karlebach 2008] Karlebach, G., & Shamir, R. (2008). Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, *9*(10), 770-780.
- Kindermann, R., & Snell, J. L. (1980). *Markov random fields and their applications* (Vol. 1). Providence, RI: American Mathematical Society.
- [Kolaczyk 2009] Kolaczyk, Eric D. *Statistical analysis of network data: methods and models*. Springer, 2009.
- [Koller 2009] Koller, Daphne, and Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.
- [Lauritzen 1996] Lauritzen, Steffen L. *Graphical models*. Oxford University Press, 1996.
- [Li 2007] Fei-Fei, L., Fergus, R., & Perona, P. (2007). Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, *106*(1), 59-70.
- [Loeliger 2007] Loeliger, H-A., Justin Dauwels, Junli Hu, Sascha Korl, Li Ping, and Frank R. Kschischang. "The factor graph approach to model-based signal processing." *Proceedings of the IEEE* 95, no. 6 (2007): 1295-1322.
- [Liu 2009] Liu, H., Lafferty, J., & Wasserman, L. (2009). The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *The Journal of Machine Learning Research*, *10*, 2295-2328.
- [Leskovec 2010] Leskovec, J., Chakrabarti, D., Kleinberg, J., Faloutsos, C., & Ghahramani, Z. (2010). Kronecker graphs: An approach to modeling networks. *The Journal of Machine Learning Research*, *11*, 985-1042.

Selected references (3)

- [Meng, Wei, Wiesel, H 2014]. Z Meng, D Wei A Wiesel, AO Hero, Distributed Learning of Gaussian Graphical Models via Marginal Likelihoods. To appear in IEEE Trans. On Signal Processing, 2014.
- [Miller 2013] Miller, Benjamin A., Nicholas Arcolano, and Nadya T. Bliss. "Efficient anomaly detection in dynamic, attributed graphs: Emerging phenomena and big data." *Intelligence and Security Informatics (ISI), 2013 IEEE International Conference on*. IEEE, 2013.
- [Nadakuditi,Newman 2012] Nadakuditi, Raj Rao, and Mark EJ Newman. "Graph spectra and the detectability of community structure in networks." *Physical review letters* 108, no. 18 (2012): 188701.
- [Nelder 1972] Nelder, J. A., & Baker, R. J. (1972). *Generalized linear models*. John Wiley & Sons.
- [Newman 2001] M. E. J. Newman, *Phys. Rev. E* **64**, 016132 (2001).
- [Newman 1999] Newman, M. E., Barkema, G. T., & Newman, M. E. J. (1999). *Monte Carlo methods in statistical physics* (Vol. 13). Oxford: Clarendon Press.
- [Ortiz-Arroyo 2010] Ortiz-Arroyo, D. (2010). Discovering sets of key players in social networks. In *Computational Social Network Analysis* (pp. 27-47). Springer London.
- [Oselio 2014] Oselio, Brandon, Alex Kulesza, and Alfred O. Hero III. "Multi-layer graph analysis for dynamic social networks.“ IEEE JSTP 2014. Availlable as *arXiv preprint arXiv:1309.5124* (2013).
- [Oselio, Kulesza, H SBP 2014] B. Oselio, J. Kulesza and A.O. Hero, "Multi-objective optimization for multi-level networks", Proc. of Workshop on Social Computing, Behavioral-Cultural Modeling and Prediction, pp. 129-136. Washington DC, April 2014.
- [Pinar 2011] Pinar, A., Seshadhri, C., & Kolda, T. G. (2011). The similarity between stochastic kronecker and chung-lu graph models. *CoRR, vol. abs/1110.4925*.
- [Rajaratnam 2008] Rajaratnam, B., Massam, H., Carvalho, C. M. (2008). Flexible covariance estimation in graphical Gaussian models. *The Annals of statistics*, *36*(6), 2818-2849.
- [Rao, Hero, States, Engel 2008] A. Rao, A. O. Hero III, D.J. States, and J.D. Engel, "Using Directed Information to Build Biologically Relevant Influence Networks," in Journal on Bioinformatics and Computational Biology, vol. 6, no.3, pp. 493-519, June 2008
- [Ravikumar 2010] Ravikumar, P., Wainwright, M. J., & Lafferty, J. D. (2010). High-dimensional Ising model selection using ℓ1-regularized logistic regression. *The Annals of Statistics*, *38*(3), 1287-1319.
- [Shih, H, 2003] MF Shih ands AO Hero, Unicast-based inference of network link delay distributions using mixed finite mixture models, *IEEE Trans. on Signal Processing*, vol. 51, No. 9, pp. 2219-2228, Aug. 2003.
- [Shmulevich 2002] Shmulevich, Ilya, et al. "Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks." *Bioinformatics* 18.2 (2002): 261-274.

Selected references (4)

- [Shuman, Narang, Frossard, Ortega, Vanderghenst 2013]Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., & Vandergheynst, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *Signal Processing Magazine, IEEE*, *30*(3), 83-98.
- [Sudderth 2010] Sudderth, E. B., Ihler, A. T., Isard, M., Freeman, W. T., & Willsky, A. S. (2010). Nonparametric belief propagation. *Communications of the ACM*, *53*(10), 95-103.
- [Steggles 2007] Steggles, L. J., Banks, R., Shaw, O., & Wipat, A. (2007). Qualitatively modelling and analysing genetic regulatory networks: a Petri net approach. *Bioinformatics*, *23*(3), 336-343.
- [Strogatz 1998] Watts, D. J.; Strogatz, S. H. (1998). "Collective dynamics of 'small-world' networks". *Nature* **393** (6684): 440–442
- [Tsilgkaridis 2013] Tsiligkaridis, T., Hero, A. O., & Zhou, S. (2013). On convergence of kronecker graphical lasso algorithms. *IEEE transactions on signal processing*, *61*(5-8), 1743-1755.
- [Wainright 2008] Wainwright, Martin J., and Michael I. Jordan. "Graphical models, exponential families, and variational inference." *Foundations and Trends® in Machine Learning* 1.1-2 (2008): 1-305.
- [Westveld 2011] A. H. Westveld and P. D. Hoff (2011). A mixed effects model for longitudinal relational and network data, with applications to international trade and conflict. *Ann. Appl. Statist.*, 5(2A):843–872.
- [Wiesel 2010] A Wiesel, YC. Eldar, and O. Hero. "Covariance estimation in decomposable Gaussian graphical models." *Signal Processing, IEEE Transactions on* 58.3 (2010): 1482-1492.
- [Wiesel, H 2009]. A Wiesel and AO Hero, Decomposable Principal Components Analysis, IEEE Trans. on Signal Processing, vol. 57, no. 11, pp. 4369-4378, 2009.
- [Watts 1999] Watts, D. Networks, Dynamics, and the Small-World Phenomenon, American Journal of Sociology, Vol. 105, No. 2 (September 1999), pp. 493-527
- [Willsky 2002] Willsky, Alan S. Multiresolution Markov models for signal and image processing. *Proceedings of the IEEE* 90.8 (2002): 1396-1458.
- [Wang, Wong 1987] Wang, Y. J., & Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, *82*(397), 8-19.
- [Yang 2012] Yang, E., Ravikumar, P. D., Allen, G. I., & Liu, Z. (2012, December). Graphical Models via Generalized Linear Models. In *NIPS* (Vol. 25, pp. 1367-1375).
- [Xu 2014] Xu, Kevin S., Mark Kliger, and Alfred O. Hero III. Adaptive evolutionary clustering. *Data Mining and Knowledge Discovery* 28.2 (2014): 304-336.
- [Xu, Kliger, H 2013] K.S. Xu, M. Kliger, AO Hero, A regularized graph layout framework for dynamic network visualization, *J. Data Mining and Knowledge Discovery*, Volume 27, Issue 1, pp 84-116, July 2013,