W-graphs and their inference

(with emphasize en variational Bayes)

S. Robin + P. Latouche, S. Ouadah

INRA / AgroParisTech

COSTNET, Ribno, September 2016

Outline

Exchangeable graphs

W-graph & graphon function

Statistical inference

Variational Bayes inference of W-graph

From graphon to residual graphon

Some notations

Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$: $\mathcal{V}$ = vertices $:= \{1, \dots n\}$, $\mathcal{E}$ = edges $\subset \mathcal{V} \times \mathcal{V}$

Adjacency matrix $Y = [Y_{ij}]$:

$$Y : n \times n, \qquad Y_{ij} = \mathbb{I}\{(i,j) \in \mathcal{E}\}$$

Random graph: defined by the joint distribution of all edges

$$p(Y) = p\left([Y_{ij}]\right)$$

Undirected graph: $Y_{ij} = Y_{ji}$

Exchangeability

Classical exchangeability: for any permutation σ

$$P([Y_i = y_i]) = P([Y_i = y_{\sigma(i)}]).$$

Joint exchangeability for two dimensional arrays [9]:

$$P([Y_{ij}] = [y_{ij}]) = P([Y_{ij}] = [y_{\sigma(i)\sigma(j)}])$$

for any permutation σ (applied to both i and j).

Exchangeability

Classical exchangeability: for any permutation σ

$$P([Y_i = y_i]) = P([Y_i = y_{\sigma(i)}]).$$

Joint exchangeability for two dimensional arrays [9]:

$$P([Y_{ij}] = [y_{ij}]) = P([Y_{ij}] = [y_{\sigma(i)\sigma(j)}])$$

for any permutation σ (applied to both i and j).

(Separate exchangeability: $P([Y_{ij}] = [y_{ij}]) = P([Y_{ij}] = [y_{\sigma(i)\tau(j)}])$
for any permutations σ and τ $\longrightarrow$ directed graphs.)

Aldous-Hoover theorem

Theorem: $Y = [Y_{ij}]$ is exchangeable iff there exists $F : [0,1]^3 \mapsto \{0,1\}$,

$$[Y_{ij}] \stackrel{d}{=} [F(U_i, U_j, U_{ij})]$$

where $(U_i)_i$ and $(U_{ij})_{i,j}$ are iid $\mathcal{U}[0,1]$.

Aldous-Hoover theorem

Theorem: $Y = [Y_{ij}]$ is exchangeable iff there exists $F : [0,1]^3 \mapsto \{0,1\}$,

$$[Y_{ij}] \stackrel{d}{=} [F(U_i, U_j, U_{ij})]$$

where $(U_i)_i$ and $(U_{ij})_{i,j}$ are iid $\mathcal{U}[0,1]$.

Properties:

- The Y_{ij}'s are not independent (because of the U_i's and U_j's)

Aldous-Hoover theorem

Theorem: $Y = [Y_{ij}]$ is exchangeable iff there exists $F : [0,1]^3 \mapsto \{0,1\}$,

$$[Y_{ij}] \stackrel{d}{=} [F(U_i, U_j, U_{ij})]$$

where $(U_i)_i$ and $(U_{ij})_{i,j}$ are iid $\mathcal{U}[0,1]$.

Properties:

- The Y_{ij}'s are not independent (because of the U_i's and U_j's)
- The Y_{ij}'s are conditionally independent given the U_i's:

$$[Y_{ij}]|(U_i = u_i)_i \stackrel{d}{=} [F(u_i, u_j, U_{ij})]$$

Some exchangeable random graphs

State-space models:

$$\begin{aligned} \{Z_i\}_i \text{ iid} &\sim \pi \\ \{Y_{ij}\}_{ij} \text{ indep.} \mid \{Z_i\}_i &\sim \mathcal{B}[\gamma(Z_i, Z_j)] \end{aligned}$$

Some exchangeable random graphs

State-space models:

$$\begin{aligned} \{Z_i\}_i \text{ iid} &\sim \pi \\ \{Y_{ij}\}_{ij} \text{ indep.} \mid \{Z_i\}_i &\sim \mathcal{B}[\gamma(Z_i, Z_j)] \end{aligned}$$

- Latent space [13]*: $Z_i \in \mathbb{R}^d, \quad \gamma(Z_i, Z_j) = f(\|Z_i - Z_j\|)$
- Model-based clustering [12]: $\pi =$ Gaussian mixture,
- Stochastic bloc-model (SBM) [14,23]: $\pi = \mathcal{M}, \quad \gamma = [\gamma_{k\ell}]$
- Continuous version of SBM [8]*: $Z_i \in \mathcal{S}^d$
- W-graph [20,9]: $\pi = \mathcal{U}_{[0,1]}$

See [21] for a statistical review + [3] for theoretical properties.

(*): Z_i not explicitly random

W-graph

- Graphon function:

$$w : [0,1]^2 \mapsto [0,1]$$

W-graph

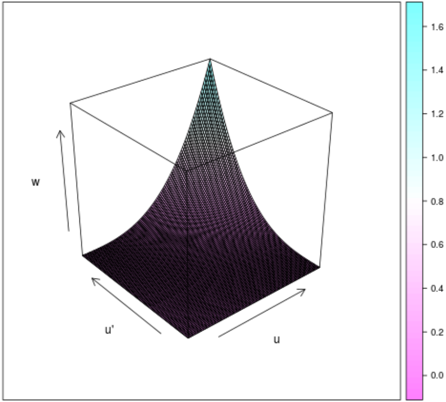
- Graphon function:

$$w : [0,1]^2 \mapsto [0,1]$$

- $\{U_i\}_i$ iid $\sim \mathcal{U}_{[0,1]}$

W-graph

Graphon function $w(u, u')$



- Graphon function:

$$w : [0,1]^2 \mapsto [0,1]$$

- $\{U_i\}_i$ iid $\sim \mathcal{U}_{[0,1]}$

- $\{Y_{ij}\}_{ij}$ indep. $\mid \{U_i\}_i$:

$$Y_{ij} | U_i, U_j \sim \mathcal{B}[w(U_i, U_j)]$$

W-graph

Graphon function $w(u, u')$

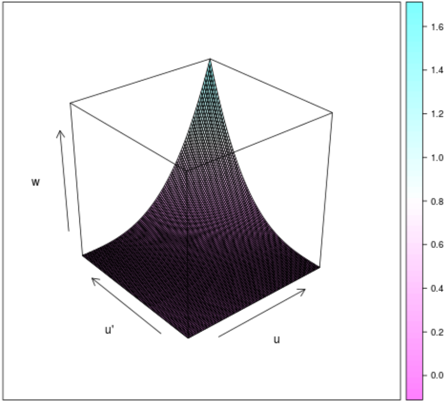
- Graphon function:

$$w : [0,1]^2 \mapsto [0,1]$$

- $\{U_i\}_i$ iid $\sim \mathcal{U}_{[0,1]}$

- $\{Y_{ij}\}_{ij}$ indep. $\mid \{U_i\}_i$:

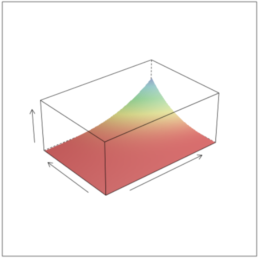
$$Y_{ij} | U_i, U_j \sim \mathcal{B}[w(U_i, U_j)]$$



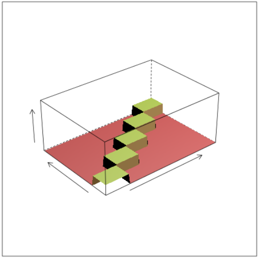
Aldous-Hoover representation: $F(U_i, U_j, U_{ij}) = \mathbb{I}\{U_{ij} < w(U_i, U_j)\}$

Some ideal graphons

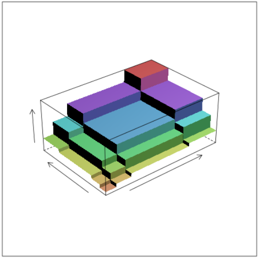
'Scale free'



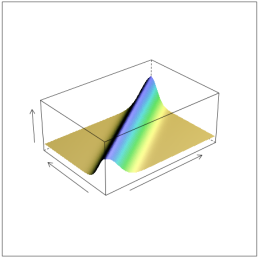
Community



SBM



Small world



W-graph as limit for dense graphs

Asymptotic framework: (a crude rephrasing of [20,9])

- $(\mathcal{G}_n)$ = sequence of exchangeable graphs with increasing size (say n)
- $t(F, G)$ = number of occurrences of subgraph F (with k nodes) in G normalized by $n_{[k]}$.

W-graph as limit for dense graphs

Asymptotic framework: (a crude rephrasing of [20,9])

- $(\mathcal{G}_n)$ = sequence of exchangeable graphs with increasing size (say n)
- $t(F, G)$ = number of occurrences of subgraph F (with k nodes) in G normalized by $n_{[k]}$.

Convergence: If for any finite set of fixed subgraphs $(F_1, \dots F_m)$, $(t(F_1, \mathcal{G}_n), \dots t(F_m, \mathcal{G}_n))$ converge in distribution,
then

- there exist a W-graph $\mathcal{G}$ such that $(\mathcal{G}_n) \xrightarrow{d} \mathcal{G}$;
- for any fixed F, $\mathbb{E}t(F, \mathcal{G}_n) \longrightarrow f(F)$ where

$$f(F) = \int_{[0,1]^k} \prod_{(i,j) \in \mathcal{E}(F)} w(u_i, u_j) \, \mathrm{d}u_1 \dots \, \mathrm{d}u_k$$

Some comments

Questions

- How strong is the condition

$$(t(F_1, \mathcal{G}_n), \dots t(F_m, \mathcal{G}_n)) \text{ converge in distribution?}$$

- Does it hold for most popular state-space models?
- Does it hold for other popular state graph models (e.g. ERGM [10])?

Some comments

Questions

- How strong is the condition

$$(t(F_1, \mathcal{G}_n), \dots t(F_m, \mathcal{G}_n)) \text{ converge in distribution?}$$

- Does it hold for most popular state-space models?
- Does it hold for other popular state graph models (e.g. ERGM [10])?

Comments: If so

- One should be able to derive the graphon of any of these models.
- Only pairwise interactions (asymptotically) matter.
- More complex patterns (e.g. triangles in ERGM) are (asymptotically) useless.

Identifiability

Obvious identifiability problem: consider $\phi : [0, 1] \mapsto [0, 1]$ measure preserving, then the two graphons

$$w(u, v) \qquad \text{and} \qquad w'(u, v) = w(\phi(u), \phi(v))$$

give raise to the same random graph [9,26].

Identifiability

Obvious identifiability problem: consider $\phi : [0,1] \mapsto [0,1]$ measure preserving, then the two graphons

$$w(u,v) \qquad \text{and} \qquad w'(u,v) = w(\phi(u), \phi(v))$$

give raise to the same random graph [9,26].

Degree function: Denoting $D_i = \sum_{j \neq i} Y_{ij}$, $\mathbb{E}(D_i | U_i = u) = (n-1) g(u)$,

$$g(u) = \int_{[0,1]} w(u,v) \, \mathrm{d}v.$$

Identifiability

Obvious identifiability problem: consider $\phi : [0,1] \mapsto [0,1]$ measure preserving, then the two graphons

$$w(u,v) \qquad \text{and} \qquad w'(u,v) = w(\phi(u), \phi(v))$$

give raise to the same random graph [9,26].

Degree function: Denoting $D_i = \sum_{j \neq i} Y_{ij}$, $\mathbb{E}(D_i | U_i = u) = (n-1) g(u)$,

$$g(u) = \int_{[0,1]} w(u,v) \, \mathrm{d}v.$$

Identifiability condition [2,26,6]:

$$g(u) \text{ strictly increasing.}$$

Outline

Exchangeable graphs

W-graph & graphon function

Statistical inference

Variational Bayes inference of W-graph

From graphon to residual graphon

Infering w

Main issues

- W-graph is a state-space for graph = incomplete data model
  $\rightarrow$ Specificity of graph models: $p([U_i]|Y)$ intractable[1].

- W-graph is a graph model
  $\rightarrow$ No prior ordering of the nodes is given in general.

[1] due to graph moralization

Infering w

Main issues

- W-graph is a state-space for graph = incomplete data model
  $\rightarrow$ Specificity of graph models: $p([U_i]|Y)$ intractable[1].
- W-graph is a graph model
  $\rightarrow$ No prior ordering of the nodes is given in general.

Specificity of W-graph

- Recent interest in the statistical community (since 2014)
  $\rightarrow$ but explosion since then.
- Main interest = flexibility of w
  $\rightarrow$ Mostly non or semi-parametric methods.

[1] due to graph moralization

Low-rank connexion probability matrix

Connectivity matrix: $P(U) = [P_{ij}] := [w(U_i, U_j)]$

Low-rank connexion probability matrix

Connectivity matrix: $P(U) = [P_{ij}] := [w(U_i, U_j)]$

Least-square estimate:

$$\widehat{P} = \arg\min_P \sum_{i,j} (Y_{ij} - P_{ij})^2 \qquad \text{s.t. } P \text{ has low rank.}$$

Low-rank connexion probability matrix

Connectivity matrix: $P(U) = [P_{ij}] := [w(U_i, U_j)]$

Least-square estimate:

$$\widehat{P} = \arg\min_P \sum_{i,j} (Y_{ij} - P_{ij})^2 \qquad \text{s.t. } P \text{ has low rank.}$$

Some references:

- [7]: thresholding the singular values of Y + bounds on the MSE
- [26]: same principle + smoothing of the resulting graphon.

SBM-based approximations

SBM = W-graph with block-wise constant graphon function.

General strategy:

1. Assign nodes to block (+ order the blocks wrt degree)
2. Estimate $w(u, v)$ with the empirical between block connectivity

[2] conditional on a correct block allocation

[3] including optimization of block allocation

SBM-based approximations

SBM = W-graph with block-wise constant graphon function.

General strategy:

1. Assign nodes to block (+ order the blocks wrt degree)
2. Estimate $w(u, v)$ with the empirical between block connectivity

A series of works:

- [24]: Constant block width h (after degree ordering) + bound on $MISE(\widehat{w})$[2]
- [6]: Same idea + smoothing between neighbor blocks
- [16,11]: Convergence rate for the least-square estimate[3] of a sparse graphon $w_n(u, v) = \rho_n w(u, v)$
- [18]: Bayesian model averaging of SBM with increasing number of blocks.
- And more [5,4]...

---

[2] conditional on a correct block allocation

[3] including optimization of block allocation

Outline

Exchangeable graphs

W-graph & graphon function

Statistical inference

Variational Bayes inference of W-graph

From graphon to residual graphon

Variational Bayes inference: General principle

$Y =$ observed data, $Z =$ latent variable, $\theta =$ parameter.

Frequentist and Bayesian inference often requires

$$p(Z|Y), \qquad p(\theta|Y) \quad \text{or} \quad p(\theta, Z|Y).$$

Variational Bayes inference: General principle

Y = observed data, Z = latent variable, θ = parameter.

Frequentist and Bayesian inference often requires

$$p(Z|Y), \qquad p(\theta|Y) \quad \text{or} \quad p(\theta, Z|Y).$$

Variational inference: find $\widetilde{p}(\cdot) \approx p(\cdot|Y)$.

Variational Bayes inference: General principle

Y = observed data, Z = latent variable, θ = parameter.

Frequentist and Bayesian inference often requires

$$p(Z|Y), \qquad p(\theta|Y) \quad \text{or} \quad p(\theta, Z|Y).$$

Variational inference: find $\widetilde{p}(\cdot) \approx p(\cdot|Y)$.

Typically [15,1,22,25]

$$\widetilde{p} = \arg\min_{q \in \mathcal{Q}} D[q(\cdot) \parallel p(\cdot|Y)]$$

- $D[q \parallel p] = KL[q \parallel p]$
- $q(\theta) = \mathcal{N}$
- $q(Z) = \prod_i q_i(Z_i)$
- $q(\theta, Z) = q(\theta)q(Z)$.

Variational Bayes inference of the graphon

SBM with K blocks:

- π block widths
- γ block heights

Variational Bayes inference of the graphon

SBM with K blocks:

- π block widths
- γ block heights

VBEM inference [17]:

$$\widetilde{p}(\pi), \ \widetilde{p}(\gamma_{k\ell}), \ \widetilde{p}(Z_i)$$

$\rightarrow \widetilde{\mathbb{E}}_K^{SBM}[w(u, v)]$

Variational Bayes inference of the graphon

SBM with K blocks:

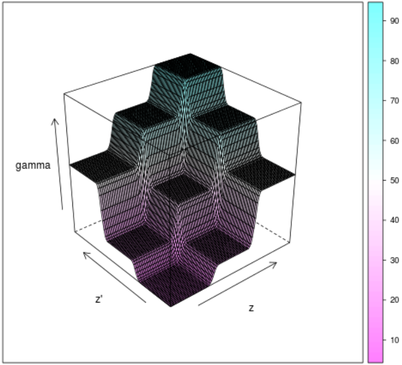
- π block widths
- γ block heights

VBEM inference [17]:

$$\widetilde{p}(\pi), \ \widetilde{p}(\gamma_{k\ell}), \ \widetilde{p}(Z_i)$$

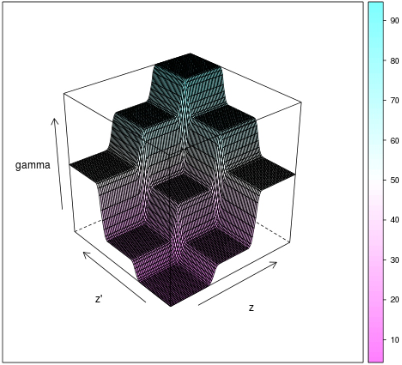
$\rightarrow \widetilde{\mathbb{E}}_K^{SBM}[w(u,v)]$

Inferred graphon with SBM_K



Variational Bayes inference of the graphon

Inferred graphon with SBM_K



SBM with K blocks:

- π block widths
- γ block heights

VBEM inference [17]:

$$\widetilde{p}(\pi), \ \widetilde{p}(\gamma_{k\ell}), \ \widetilde{p}(Z_i)$$

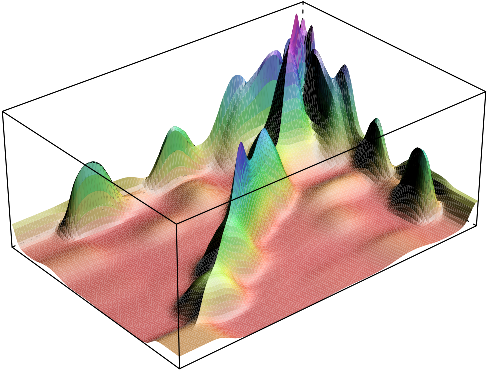
$\rightarrow \widetilde{\mathbb{E}}_K^{SBM}[w(u,v)]$

Bayesian model averaging [18]:

$$\widetilde{\mathbb{E}}[w(u,v)] = \sum_K \widetilde{P}(K)\widetilde{\mathbb{E}}_K^{SBM}[w(u,v)]$$

How to interpret a graphon?

French political blogs: $n = 196$ nodes [18]



$\rightarrow$ Depicts the heterogeneity of the network.

Outline

Exchangeable graphs

W-graph & graphon function

Statistical inference

Variational Bayes inference of W-graph

From graphon to residual graphon

Accounting for covariates [19]

Data: Y = observed (binary) network, x = (edge) covariates

Accounting for covariates [19]

Data: Y = observed (binary) network, x = (edge) covariates

Questions:

- Does x explain the topology of Y?
- Residual (wrt x) heterogeneity in Y?

Accounting for covariates [19]

Data: Y = observed (binary) network, x = (edge) covariates

Questions:

- Does x explain the topology of Y?
- Residual (wrt x) heterogeneity in Y?

Logistic regression (H_0): $\text{logit } P(Y_{ij} = 1) = \beta_0 + x_{ij}^\intercal \beta$

Accounting for covariates [19]

Data: Y = observed (binary) network, x = (edge) covariates

Questions:

- Does x explain the topology of Y?
- Residual (wrt x) heterogeneity in Y?

Logistic regression (H_0): logit $P(Y_{ij} = 1) = \beta_0 + x_{ij}^\intercal \beta$

Logistic regression + graphon residual term: (U_i) iid $\sim \mathcal{U}[0, 1]$,

$$\text{logit } P(Y_{ij} = 1 | U_i, U_j) = \omega(U_i, U_j) + x_{ij}^\intercal \beta$$

Accounting for covariates [19]

Data: Y = observed (binary) network, x = (edge) covariates

Questions:

- Does x explain the topology of Y?
- Residual (wrt x) heterogeneity in Y?

Logistic regression (H_0): logit $P(Y_{ij} = 1) = \beta_0 + x_{ij}^\intercal \beta$

Logistic regression + graphon residual term: (U_i) iid $\sim \mathcal{U}[0, 1]$,

$$\text{logit } P(Y_{ij} = 1 | U_i, U_j) = \omega(U_i, U_j) + x_{ij}^\intercal \beta$$

Goodness of fit (GOF): Check if $\omega(u, v) = \text{cst} \qquad (= \beta_0)$

Goodness-of-fit as model comparison

Auxiliary model M_K:

$$\text{logit } P(Y_{ij} = 1 | U_i, U_j, K) = \omega_K^{SBM}(U_i, U_j) + x_{ij}^\intercal \beta.$$

Goodness-of-fit as model comparison

Auxiliary model M_K:

$$\text{logit } P(Y_{ij} = 1 | U_i, U_j, K) = \omega_K^{SBM}(U_i, U_j) + x_{ij}^\intercal \beta.$$

Goodness of fit:

$$\begin{array}{lllll} H_0 & = & \{\text{logistic regression is sufficient}\} & = & M_1 \\ H_1 & = & \{\text{logistic regression is not sufficient}\} & = & \bigcup_{K>1} M_K \end{array}$$

GOF is a assessed if

$$P(H_0|Y) = P(M_1|Y) \text{ is large.}$$

Goodness-of-fit as model comparison

Auxiliary model M_K:

$$\text{logit } P(Y_{ij} = 1 | U_i, U_j, K) = \omega_K^{SBM}(U_i, U_j) + x_{ij}^\intercal \beta.$$

Goodness of fit:

$$\begin{array}{rcccl} H_0 & = & \{\text{logistic regression is sufficient}\} & = & M_1 \\ H_1 & = & \{\text{logistic regression is not sufficient}\} & = & \bigcup_{K>1} M_K \end{array}$$

GOF is a assessed if

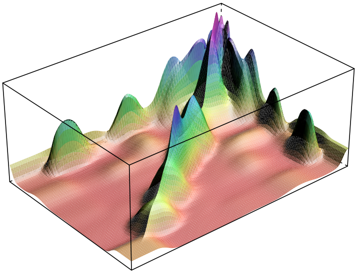
$$P(H_0|Y) = P(M_1|Y) \text{ is large.}$$

Actually: use $\widetilde{P}(H_0)$ and $\widetilde{P}(M_1)$.

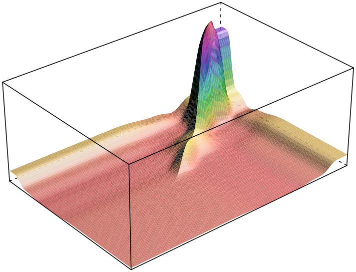
Political blog network

$n = 196$ blogs ($N = 19110$ pairs), 3 covariates, density $= .075$

Inferred graphon (no covariate)



Residual graphon (3 covariates)

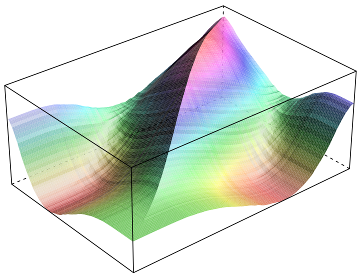


$\widetilde{P}(H_0) \simeq 10^{-172}$

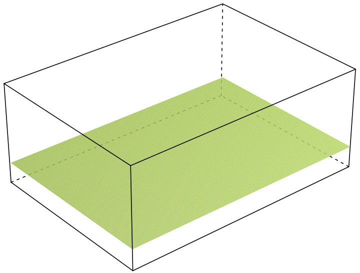
Florentine business

$n = 16$ families ($N = 120$ pairs), 3 covariates, density $= .12$

Inferred graphon (no covariate)



Residual graphon (3 covariates)



$\widetilde{P}(H_0) = .991$

Some more examples

network	size (n)	nb. covariates (d)	density	$\hat{p}(H_0 \mid Y)$
Blog	196	3	0.075	3e-172
Tree	51	3	0.54	2e-115
Karate	34	8	0.14	3e-2
Florentine (marriage)	16	3	0.17	0.995
Florentine (business)	16	3	0.125	0.991
Faux Dixon High	248	17	0.02	1
CKM	219	39	0.015	1
AddHealth 67	530	21	0.007	2e-25

Conclusion

W-graph

- A well established object in the probability literature
- A more recent interest in the statistical community, with several theoretical results
- Still not very popular among practitioners

Conclusion

W-graph

- A well established object in the probability literature
- A more recent interest in the statistical community, with several theoretical results
- Still not very popular among practitioners

Variational Bayes inference and GOF

- R package on `github.com/platouche/gofNetwork` (soon on CRAN)
- Strongly relies on variational Bayes approximation of the posteriors.
  $\longrightarrow$ VBEM asymptotically accurate for logistic regression and SBM
  $\longrightarrow$ No clue about accuracy in the combined model.

References I

J. Beal, M. and Z. Ghahramani.
The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures.
Bayes. Statist., 7:543–52, 2003.

P. J. Bickel and A. Chen.
A nonparametric view of network models and Newman-Girvan and other modularities.
Proc. Natl. Acad. Sci. U.S.A., 106:21068–21073, Dec 2009.

B. Bollobás, S. Janson, and O. Riordan.
The phase transition in inhomogeneous random graphs.
Rand. Struct. Algo., 31(1):3–122, 2007.

C. Borgs, J.T. Chayes, H. Cohn, and S. Ganguly.
Consistent nonparametric estimation for heavy-tailed sparse graphs.
arXiv preprint arXiv:1508.06675v2, 2016.

D. Cai, N. Ackerman, and C. Freer.
An iterative step-function estimator for graphons.
arXiv preprint arXiv:1412.2129v2, 2016.

S. H. Chan and E. Airoldi.
A consistent histogram estimator for exchangeable graph models.
In *ICML*, pages 208–216, 2014.

S. Chatterjee.
Matrix estimation by universal singular value thresholding.
The Annals of Statistics, 43(1):177–214, 2015.

J.-J. Daudin, L. Pierre, and C. Vacher.
Model for heterogeneous random networks using continuous latent variables and an application to a tree–fungus network.
Biometrics, 66(4):1043–1051, 2010.

References II

P. Diaconis and S. Janson.
Graph limits and exchangeable random graphs.
Rend. Mat. Appl., 7(28):33–61, 2008.

O. Frank and D. Strauss.
Markov graphs.
J. Amer. Statist. Assoc., 81(395):832–842, 1986.

C. Gao, Y. Lu, and H.H. Zhou.
Rate-optimal graphon estimation.
The Annals of Statistics, 43(6):2624–2652, 2015.

M. S. Handcock, A. E. Raftery, and J. M. Tantrum.
Model-based clustering for social networks.
Journal of the Royal Statistical Society: Series A (Statistics in Society), 170(2):301–354, 2007.
doi: 10.1111/j.1467-985X.2007.00471.x.

P. D. Hoff, A. E. Raftery, and M. S. Handcock.
Latent space approaches to social network analysis.
J. Amer. Statist. Assoc., 97(460):1090–98, 2002.

P.W. Holland and S. Leinhardt.
Structural sociometry.
Perspectives on social network research, pages 63–83, 1979.

T. Jaakkola.
Advanced mean field methods: theory and practice, chapter Tutorial on variational approximation methods.
MIT Press, 2000.

O. Klopp, A.B. Tsybakov, and N. Verzelen.
Oracle inequalities for network models and sparse graphon estimation.
arXiv preprint arXiv:1507.04118v2, 2016.

References III

P. Latouche, E. Birmelé, and C. Ambroise.
Variational bayesian inference and complexity control for stochastic block models.
Statis. Model., 12(1):93–115, 2012.

P. Latouche and S. Robin.
Variational bayes model averaging for graphon functions and motif frequencies inference in W-graph models.
Statistics and Computing, pages 1–13, 2015.

P. Latouche, S. Robin, and S. Ouadah.
Goodness of fit of logistic models for random graphs.
Technical report, arXiv:1508.00286, 2015.

L. Lovász and B. Szegedy.
Limits of dense graph sequences.
Journal of Combinatorial Theory, Series B, 96(6):933 – 957, 2006.

C. Matias and S. Robin.
Modeling heterogeneity in random graphs through latent space models: a selective review.
ESAIM: Proc., 47:55–74, 2014.

Tom Minka.
Divergence measures and message passing.
Technical Report MSR-TR-2005-173, Microsoft Research Ltd, 2005.

K. Nowicki and T.A.B. Snijders.
Estimation and prediction for stochastic block-structures.
J. Amer. Statist. Assoc., 96:1077–87, 2001.

S.C. Olhede and P.J. Wolfe.
Network histograms and universality of blockmodel approximation.
Proceedings of the National Academy of Sciences, 111(41):14722–14727, 2014.

References IV

M. J. Wainwright and M. I. Jordan.
Graphical models, exponential families, and variational inference.
Found. Trends Mach. Learn., 1(1–2):1–305, 2008.

J. Yang, C. Han, and E. Airoldi.
Nonparametric estimation and testing of exchangeable graph models.
In *AISTATS*, pages 1060–1067, 2014.