NETWORKS 2021

11005. Network Models and Graphical Models: A Survey

Momin M. Malik

Tuesday, 6 July 2021

Networks 2021, Session S65. Machine Learning

Scope

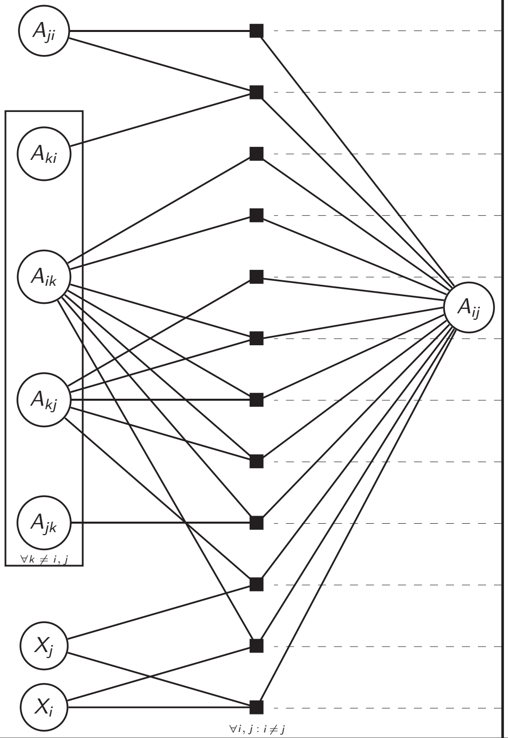
- Network models represent dependencies with graphs; graphical models (one type of which are "Bayesian networks") represent dependencies with graphs. That causes confusion. But they are very different!
- Graphical models haven't done the best job at networks, and networks haven't made use of graphical models
- This talk is primarily to clarify (or create!) a *conceptual* connection between these two types of models
- For examples of *applications* of graphical modeling to networks, see, e.g., Farasat et al. (2015), Maier et al. (2014), and Airoldi et al. (2008)

Key points

- **Graphical models** represent *dependencies* (and causal relationships) *between variables*
- **Networks models** are models of dyads, which represent *dependencies between observations*
  - Dyads can be modeled as random variables (e.g., Bernoulli for unweighted; Poisson for count; etc.)
  - Dyads are themselves dependent! (reciprocity, triadic closure, degree constraints)
  - Graphical models can represent these “dyadic dependencies”!

Key visualization

Terms: Snijders et al. (2006). From (old) joint work with Antonis Manousis and Naji Shajarisales.

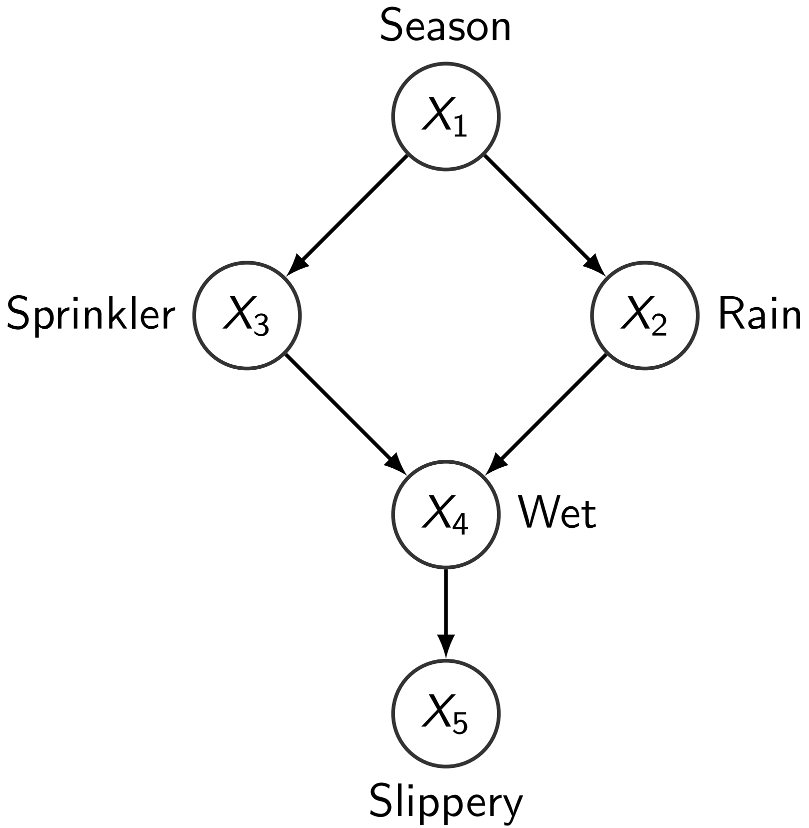


Factor graph	Parameter name	Network Motif	Parameterization	Matrix notation
	mutual dyads		$\sum_{i<j} A_{ij}A_{ji}$	$\frac{1}{2}\,\mathrm{tr}\left(\mathbf{AA}^T\right)$
	in-two-stars		$\sum_{(i,j,k)} A_{ji}A_{ki}$	$\mathrm{sum}\left(\mathbf{AA}^T\right) - \mathrm{tr}\left(\mathbf{AA}^T\right)$
	out-two-stars		$\sum_{(i,j,k)} A_{ij}A_{ik}$	$\mathrm{sum}\left(\mathbf{A}^T\mathbf{A}\right) - \mathrm{tr}\left(\mathbf{A}^T\mathbf{A}\right)$
	geom. weighted out-degrees	—	$\sum_i \exp\left\{-\alpha \sum_k A_{ik}\right\}$	$\mathrm{sum}\left(\exp\{-\alpha\ \mathrm{rowsum}\left(\mathbf{A}\right)\}\right)$
	geom. weighted in-degrees	—	$\sum_j \exp\left\{-\alpha \sum_k A_{kj}\right\}$	$\mathrm{sum}\left(\exp\{-\alpha\ \mathrm{colsum}\left(\mathbf{A}\right)\}\right)$
	alternating transitive k-triplets		$\lambda \sum_{i,j} A_{ij} \left\{1-\left(1-\frac{1}{\lambda}\right)^{\sum_{k\neq i,j} A_{ik}A_{kj}}\right\}$	$\lambda\ \mathrm{sum}\left(\mathbf{A}\odot\left(1-\left(1-\frac{1}{\lambda}\right)^{\mathbf{AA}-\mathrm{diag}(\mathbf{AA})}\right)\right)$
	alternating indep. two-paths		$\lambda \sum_{i,j} \left\{1-\left(1-\frac{1}{\lambda}\right)^{\sum_{k\neq i,j} A_{ik}A_{kj}}\right\}$	$\lambda\ \mathrm{sum}\left(1-\left(1-\frac{1}{\lambda}\right)^{\mathbf{AA}-\mathrm{diag}(\mathbf{AA})}\right)$
	two-paths (mixed two-stars)		$\sum_{(i,k,j)} A_{ik}A_{kj}$	$\mathrm{sum}\left(\mathbf{AA}\right) - \mathrm{tr}\left(\mathbf{AA}\right)$
	transitive triads		$\sum_{(i,j,k)} A_{ij}A_{jk}A_{ik}$	$\mathrm{tr}\left(\mathbf{AAA}^T\right)$
	activity effect		$\sum_i X_i \sum_j A_{ij}$	$\mathrm{sum}\left(\mathbf{X}\odot\mathrm{rowsum}\left(\mathbf{A}\right)\right)$
	popularity effect		$\sum_j X_j \sum_i A_{ij}$	$\mathrm{sum}\left(\mathbf{X}\odot\mathrm{colsum}\left(\mathbf{A}\right)\right)$
	similarity effect		$\sum_{i,j} A_{ij}\left(1-\frac{\lvert X_i-X_j\rvert}{\max_{k,l}\lvert X_k-X_l\rvert}\right)$	$\mathrm{sum}\left(\mathbf{A}\odot\mathbf{S}\right)$

What are graphical models?

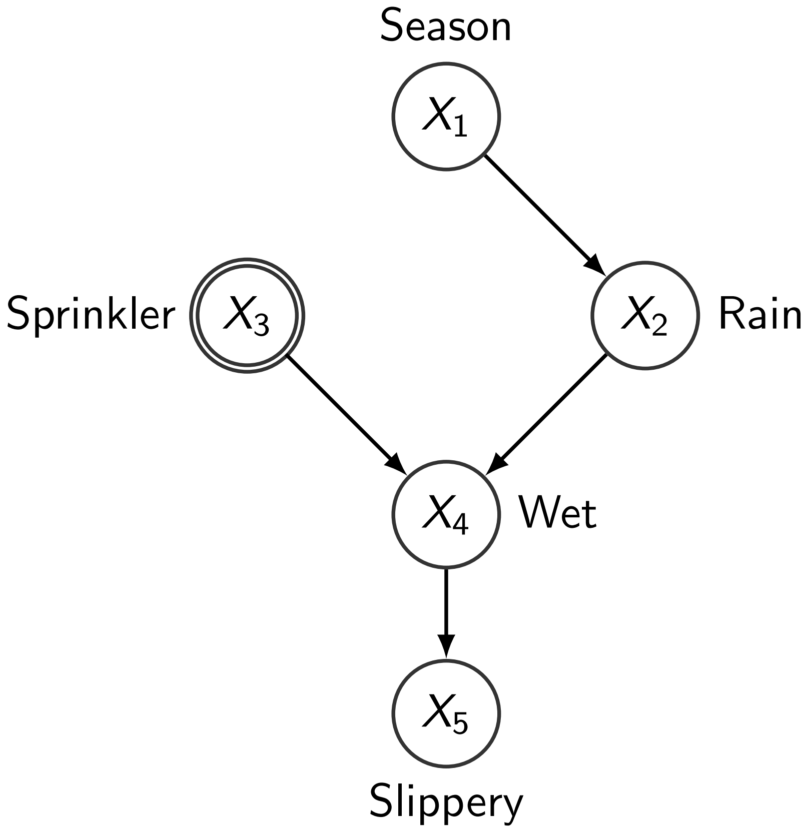
Graphical models = graphs for variables

- Like path diagrams in psych, but more formal. From CS in 90s
- Represent relationships between variables; can reason through dependencies
  - Sprinklers are not directly dependent on Rain, but if we know the grass is wet, we know *either* it rained *or* sprinklers were on (at least one is true)
- With probability distributions on the nodes, they represent *conditional independencies*
  - Equivalent to structural equation modeling (SEMs)!



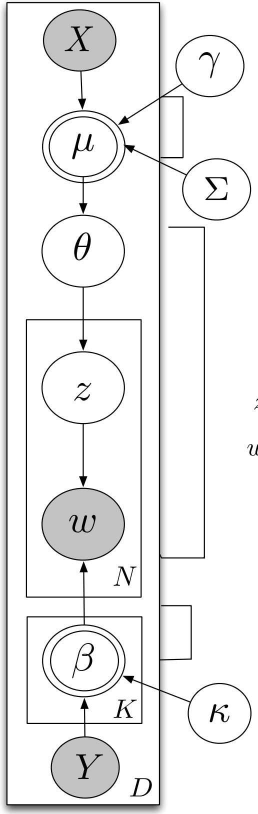
Can represent causality

- Can also reason about *causality*
- Interventions block "paths"
  - Pearl introduced the "do" operator to notate this algebraically
- Are algorithms to determine identifiability of parameters from a given (or assumed) causal structure (Bayes ball)
  - Causal *inference* techniques to estimate a causal graph (e.g., TETRAD algorithm) also exist, but theoretical guarantees for these procedures require strong, untestable, and almost certainly false assumptions



Mostly used for bookkeeping

- Most machine learning applications are effectively "bookkeeping"
- E.g., in structural topic modeling (Roberts et al., 2013), a topic doesn't "cause" a document, but representing it as a directed tie is to help keep track of things for estimation



Topic Prevalence:

$$\mu_{d,k} = X_d \gamma_k$$
$$\gamma_k \sim \mathcal{N}(0, \sigma_k^2)$$
$$\sigma_k^2 \sim \text{Gamma}(s^\gamma, r^\gamma)$$

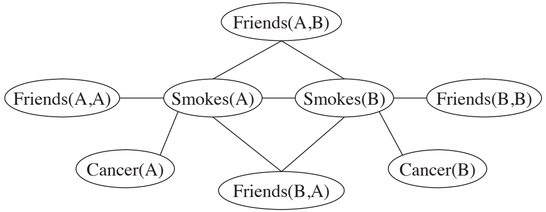
Language Model:

$$\theta_d \sim \text{LogisticNormal}(\mu_d, \Sigma)$$
$$z_{d,n} \sim \text{Mult}(\theta_d)$$
$$w_{d,n} \sim \text{Mult}(\beta_d^{k=z_{d,n}})$$

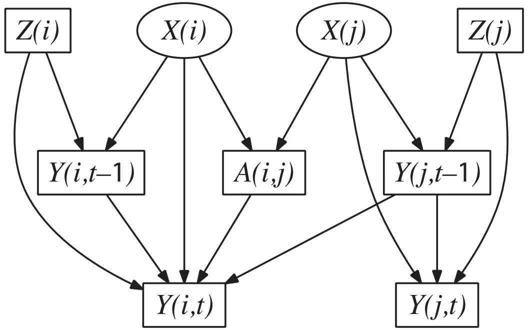
Topical Content:

$$\beta_{d,v}^k \propto \exp(m_v + \kappa_v^{\cdot,k} + \kappa_v^{y,\cdot} + \kappa_v^{y,k})$$
$$\kappa_v^{y,k} \sim \text{Laplace}(0, \tau_v^{y,k})$$
$$\tau_v^{y,k} \sim \text{Gamma}(s^\kappa, r^\kappa)$$

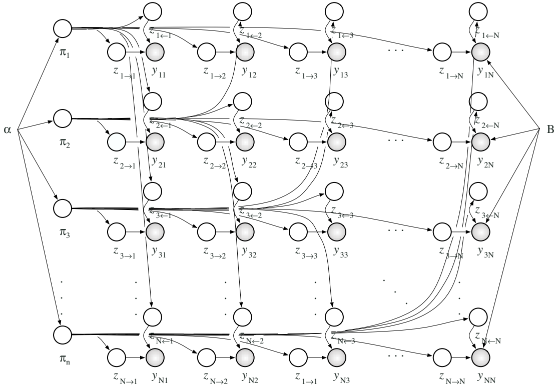
Networks in graphical models



Network edges as graphical model nodes in a "relational Markov network" (Getoor & Taskar, 2007): awkward



Network is only a confounder (Shalizi & Thomas, 2011), no direct dependencies between edges



Full representation, but only of a single type of dependency (block membership) (Airoldi et al., 2008)

Two "directions" of dependencies

- If covariates are independent (trivial graphical model, what we usually assume in linear regression), the joint distribution of a response and a design matrix is:

$$p(Y, \mathbf{X}) = p(Y, X_1, \ldots, X_d) = p(Y | X_1, \ldots, X_d) \prod_{j=1}^{d} p(X_j)$$

- (If we assume fixed $\mathbf{X}$, as we usually do, the probabilities of X_j's go away)
- But a true, complete joint factorization of the conditional distribution would be over *observations* as well:

$$p(Y | \mathbf{X}) = p(y_1, \ldots, y_n | x_{11}, \ldots, x_{1d}, x_{21}, \ldots, x_{2d}, \ldots\ldots, x_{n1}, \ldots, x_{nd})$$

- An iid assumption applies to the *observations*, and are how we even have multiple observations to estimate anything. This looks like:

$$p(Y | \mathbf{X}) = p(y_1, \ldots, y_n | \mathbf{x}_1, \ldots, \mathbf{x}_n) \stackrel{\text{iid}}{=} \prod_{i=1}^{n} p(y_i | \mathbf{x}_i)$$

Parameterizing network models

Networks: Dependencies as observations

Better way of looking at networks: make dyads the observations

- The response is now an edge, or edge attribute
- Transform all node covariates into edge covariates, e.g.,
  - As a difference between continuous node attributes
  - Indicator for if nodes in same category or not (or, make new categories out of possible pairs, e.g., M→M, M→F, F→M, F→F)
  - As "sender" and/or "receiver" attributes

	Y	X_1	X_2	$\cdots$	X_d
1	y_1	x_{11}	x_{12}	$\cdots$	x_{1d}
2	y_2	x_{21}	x_{22}	$\cdots$	x_{2d}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	y_n	x_{n1}	x_{n2}	$\cdots$	x_{nd}

index	*from*	*to*	Y	W_1	W_2	W_3	$\cdots$
e_1	1	2	y_{12}	$\mathbf{1}(x_{11} = x_{21})$	$x_{12} - x_{22}$	x_{13}	$\cdots$
e_2	2	3	y_{23}	$\mathbf{1}(x_{11} = x_{31})$	$x_{12} - x_{32}$	x_{13}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
e_{n+1}	2	1	y_{21}	$\mathbf{1}(x_{21} = x_{11})$	$x_{22} - x_{12}$	x_{23}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$e_{2\binom{n}{2}}$	$n-1$	n	$y_{(n-1)n}$	$\mathbf{1}(x_{(n-1)1} = x_{n1})$	$x_{(n-1)2} - x_{n2}$	$x_{(n-1)3}$	$\cdots$

Parameterizing

- Descriptively,

$$a_{ij} = \begin{cases} 1 \text{ if there is a tie } i \to j \\ 0 \text{ otherwise.} \end{cases}$$

- Turn this into a random variable:

$$A_{ij} \overset{\text{iid}}{\sim} \text{Bernoulli}\,(p)$$

As a logistic regression

- Add some covariates and it can become a logistic regression:

$$A_{ij} | x_i, x_j \sim \text{Bernoulli}\left(f(x_i, x_j)^T \boldsymbol{\beta}\right)$$

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \prod_{j \neq i} f(x_i, x_j)^T \boldsymbol{\beta}^{a_{ij}} \left(f(x_i, x_j)^T \boldsymbol{\beta}\right)^{1-a_{ij}}$$

- The MLE of an intercept-only model is just the density.

$$\widehat{\beta}_{MLE} = \frac{1}{2 \times \binom{n}{2}} \sum_{i \neq j} a_{ij}$$

Problem: Dyads are dependent, too

- In the language of ERGMs, "dyadic dependencies"
  - Social networks: reciprocity makes $A_{ij} \not\perp\!\!\!\perp A_{ji}$
- The p_1 model (Holland & Leinhardt, 1981) deals with reciprocity as a one-off dependency by modeling edges as multinomial, with a cross term:

$$\begin{aligned}\mathbb{P}(A_{ij} = a_{ij}, A_{ji} = a_{ji}) = \frac{1}{k_{ij}} \exp \{\, & a_{ij}(\mu + \alpha_i + \beta_i) \\ & + a_{ji}(\mu + \alpha_j + \beta_j) \\ & + \rho a_{ij} a_{ji} \,\}\end{aligned}$$

Models for dyadic dependencies

- Stochastic blockmodels (Wang & Wong, 1987): alternative to p_1, two-level hierarchical version of Bernoulli model
- Latent space models (Hoff et al., 2002): can be seen as graphical models with observable nodes for edges, produced from hidden nodes representing latent position

Markov property for network edges

- Landmark work: Frank & Strauss (1986)
- Markov dependence assumption: "A graph is said to be a Markov graph if only incident dyads can be conditionally dependent."
- In retrospect, we can clarify this in terms of graphical models
  - The "graph" is the network, and the Markov property is of the graphical model of the network edges as Bernoulli variables

Markov property for network edges

- Remarkably, using Hammersley-Clifford, Frank & Strauss proved that the graphical model of an undirected network is Markov if and only if

$$P_\theta(\mathbf{A}) = \frac{1}{\kappa(\boldsymbol{\theta})} \exp\left\{ \theta_0 L(\mathbf{A}) + \sum_{k=1}^{n-1} \theta_k S_k(\mathbf{A}) + \theta_\tau T(\mathbf{A}) \right\}$$

Normalization constant: need to sum over $2^{2\times\binom{n}{2}}$ possible networks for each candidate $\boldsymbol{\theta}$

Non-maximal k-stars

Number of triangles

- Especially surprising part (Kolaczyk, 2009): how did triangles come out of this as a sufficient statistic??

Graphical models for networks

General modeling of dyadic dependencies

- This eventually led to Exponential-family Random Graph Models, which can model generic dependencies between edges
- Can add any sufficient statistic, although they can be collinear. E.g., two-paths are collinear with in-degrees, out-degrees, and mutual dyads (Snijders et al., 2006)

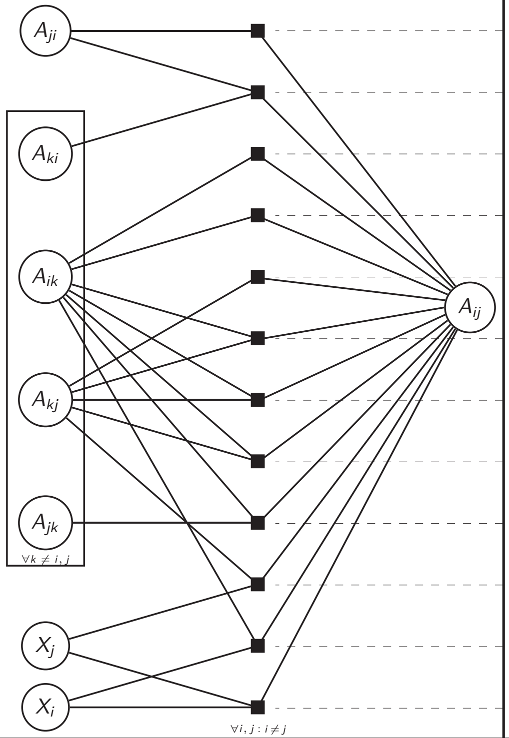
$$\sum_{i,j,k:k\neq i} A_{ij}A_{jk} = \sum_{j=1}^{n}\sum_{i,k:k\neq i} A_{ij}A_{jk} = \sum_{j=1}^{n}\left(A_{+j}A_{j+} - \sum_{i=1}^{n} A_{ij}A_{ji}\right)$$

- But: bad theoretical properties, tricky to estimate, and tricky to specify

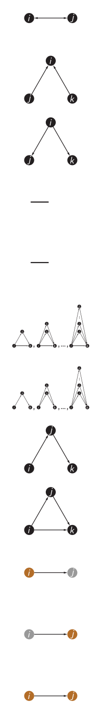
Dyadic dependencies in a graph

Terms: Snijders et al. (2006). From (old) joint work with Antonis Manousis and Naji Shajarisales.

Factor graph



Network Motif



Parameter name	Network Motif	Parameterization	Matrix notation
mutual dyads		$\sum_{i<j} A_{ij}A_{ji}$	$\frac{1}{2}\,\mathrm{tr}\left(\mathbf{AA}^T\right)$
in-two-stars		$\sum_{(i,j,k)} A_{ji}A_{ki}$	$\mathrm{sum}\left(\mathbf{AA}^T\right) - \mathrm{tr}\left(\mathbf{AA}^T\right)$
out-two-stars		$\sum_{(i,j,k)} A_{ij}A_{ik}$	$\mathrm{sum}\left(\mathbf{A}^T\mathbf{A}\right) - \mathrm{tr}\left(\mathbf{A}^T\mathbf{A}\right)$
geom. weighted out-degrees	—	$\sum_i \exp\left\{-\alpha \sum_k A_{ik}\right\}$	$\mathrm{sum}\left(\exp\{-\alpha\,\mathrm{rowsum}\left(\mathbf{A}\right)\}\right)$
geom. weighted in-degrees	—	$\sum_j \exp\left\{-\alpha \sum_k A_{kj}\right\}$	$\mathrm{sum}\left(\exp\{-\alpha\,\mathrm{colsum}\left(\mathbf{A}\right)\}\right)$
alternating transitive k-triplets		$\lambda \sum_{i,j} A_{ij}\left\{1-\left(1-\frac{1}{\lambda}\right)^{\sum_{k\neq i,j} A_{ik}A_{kj}}\right\}$	$\lambda\,\mathrm{sum}\left(\mathbf{A}\odot\left(1-\left(1-\frac{1}{\lambda}\right)^{\mathbf{AA}-\mathrm{diag}(\mathbf{AA})}\right)\right)$
alternating indep. two-paths		$\lambda \sum_{i,j}\left\{1-\left(1-\frac{1}{\lambda}\right)^{\sum_{k\neq i,j} A_{ik}A_{kj}}\right\}$	$\lambda\,\mathrm{sum}\left(1-\left(1-\frac{1}{\lambda}\right)^{\mathbf{AA}-\mathrm{diag}(\mathbf{AA})}\right)$
two-paths (mixed two-stars)		$\sum_{(i,k,j)} A_{ik}A_{kj}$	$\mathrm{sum}\left(\mathbf{AA}\right) - \mathrm{tr}\left(\mathbf{AA}\right)$
transitive triads		$\sum_{(i,j,k)} A_{ij}A_{jk}A_{ik}$	$\mathrm{tr}\left(\mathbf{AAA}^T\right)$
activity effect		$\sum_i X_i \sum_j A_{ij}$	$\mathrm{sum}\left(\mathbf{X}\odot\mathrm{rowsum}\left(\mathbf{A}\right)\right)$
popularity effect		$\sum_j X_j \sum_i A_{ij}$	$\mathrm{sum}\left(\mathbf{X}\odot\mathrm{colsum}\left(\mathbf{A}\right)\right)$
similarity effect		$\sum_{i,j} A_{ij}\left(1-\frac{\lvert X_i - X_j\rvert}{\max_{k,l}\lvert X_k - X_l\rvert}\right)$	$\mathrm{sum}\left(\mathbf{A}\odot\mathbf{S}\right)$

Conclusions

Key points, redux

- **Graphical models** represent *dependencies* (and causal relationships) *between variables*
- **Networks models** are models of dyads, which represent *dependencies between observations*
- They are not the same thing, but we can represent dyadic dependencies (dependencies between edges of a network, in processes like reciprocity and transitivity) as graphical models

Why should we care?

- Graphical models are network models are both powerful for representing, reasoning through, and modeling dependencies
  - But graphical models haven't done the best job at networks, and networks haven't made use of graphical models
  - Maybe because they are used by different communities, and the same words ("networks", "dependencies") mean subtly different things
- Clarifying the relationship of these two types of models helps head off confusion, as well as deepen our appreciation of the idea of "dependencies"
  - As well as helping to train students
- Practically: Can graphical models help create, and estimate, new statistical network models, and unify existing ones? Almost certainly, although in many cases the estimation will still be MCMC (maybe variational inference; Celisse et al., 2012)

NETWORKS 2021

Scope

Key points

What are graphical models?

Parameterizing network models

Graphical models for networks

Conclusions

References

References

Airoldi, Edoardo M., David M. Blei, Stephen E. Fienberg, and Eric P. Xing. 2008. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research* 9, 1981–2014.

Celisse, Alain, Jean-Jacques Daudin, and Laurent Pierre. 2012. Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electronic Journal of Statistics* 6, 1847–1899. doi: 10.1214/12-EJS729.

Farasat, Alireza, Alexander Nikolaev, Sargur N. Srihari, and Rachael Hageman Blair. 2015. Probabilistic graphical models in modern social network analysis. *Social Network Analysis and Mining* 5 (62). doi: 10.1007/s13278-015-0289-6

Fienberg, Stephen E. 2012, April. Graphical models for network data. Presentation at the Fields Institute Workshop on Graphical Models.

Frank, Ove and David Strauss. 1986. Markov graphs. *Journal of the American Statistical Association* 81 (395), 832–842. doi: 10.1080/01621459.1986.10478342

Getoor, Lise and Ben Taskar. 2007. *Introduction to statistical relational learning*. The MIT Press.

Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock. 2002. Latent space approaches to social network analysis. *Journal of the American Statistical Association* 97 (460), 1090–1098. doi: 10.1198/016214502388618906.

Kolaczyk, Eric D. 2009. *Statistical analysis of network data: Methods and models*. Springer. doi: 10.1007/978-0-387-88146-1.

Koller, Daphne and Nir Friedman. 2009. *Probabilistic graphical models: Principles and techniques*. The MIT Press, 2009.

Maier, Marc, Katerina Marazopoulou, and David Jensen. Reasoning about independence in probabilistic models of relational data. arXiv:1302.4381v3

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, and Edoardo M. Airoldi. 2013. The structural topic model and applied social science. Prepared for the NeurIPS 2013 Workshop on Topic Models: Computation, Application, and Evaluation.

Snijders, Tom A. B., Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock. 2006. New specifications for Exponential Random Graph Models. *Sociological Methodology* 36, 99–153. doi: 10.1111/j.1467-9531.2006.00176.x

Wang, Yuchung J. and George Y. Wong. 1987. Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association* 82 (397), 8–19. doi: 10.1080/01621459.1987. 10478385