Models for Network Graphs

Gonzalo Mateos
Dept. of ECE and Goergen Institute for Data Science
University of Rochester
gmateosb@ece.rochester.edu
http://www.hajim.rochester.edu/ece/sites/gmateos/

March 27, 2023

Random graph models

Random graph models

Small-world models

Network-growth models

Exponential random graph models

Latent network models

Random dot product graphs

Why statistical graph modeling?

- Statistical graph models are used for a variety of reasons:

1) Mechanisms explaining properties observed on real-world networks
   Ex: small-world effects, power-law degree distributions
2) Testing for 'significance' of a characteristic $\eta(G)$ in a network graph
   Ex: is the observed average degree unusual or anomalous?
3) Assessment of factors potentially predictive of relational ties
   Ex: are there reciprocity or transitivity effects in play?

- Focus today on construction and use of models for network data

Modeling network graphs

- **Def:** A model for a network graph is a collection

$$\{P_\theta(G), G \in \mathcal{G} : \theta \in \Theta\}$$

  - $\mathcal{G}$ is an ensemble of possible graphs
  - $P_\theta(\cdot)$ is a probability distribution on $\mathcal{G}$ (often write $P(\cdot)$)
  - Parameters θ ranging over values in parameter space Θ

- Richness of models derives from how we specify $P_\theta(\cdot)$

  $\Rightarrow$ Methods range from the simple to the complex

Model specification

1) Let $\mathbb{P}(\cdot)$ be uniform on $\mathcal{G}$, add structural constraints to $\mathcal{G}$

   Ex: Erdös-Renyi random graphs, generalized random graph models

2) Induce $\mathbb{P}(\cdot)$ via application of simple generative mechanisms

   Ex: small world, preferential attachment, copying models

3) Model structural features and their effect on G's topology

   Ex: exponential random graph models

4) Model propensity towards establishing links via latent variables

   Ex: stochastic block models, graphons, random dot product graphs

- Computational cost of associated inference algorithms relevant

Classical random graph models

- Assign equal probability on all undirected graphs of given order and size
  - Specify collection $\mathcal{G}_{N_v,N_e}$ of graphs $G(V,E)$ with $|V| = N_v$, $|E| = N_e$
  - Assign $\mathrm{P}(G) = \binom{N}{N_e}^{-1}$ to each $G \in \mathcal{G}_{N_v,N_e}$, where $N = |V^{(2)}| = \binom{N_v}{2}$
- Most common variant is the Erdös-Renyi random graph model $G_{n,p}$

  $\Rightarrow$ Undirected graph on $N_v = n$ vertices

  $\Rightarrow$ Edge (u,v) present w.p. p, independent of other edges
- Simulation: simply draw $N = \binom{N_v}{2} \approx N_v^2/2$ i.i.d. Ber(p) RVs
  - Inefficient when $p \sim N_v^{-1} \Rightarrow$ sparse graph, most draws are 0
  - Skip non-edges drawing Geo(p) i.i.d. RVs, runs in $O(N_v + N_e)$ time

Properties of $G_{n,p}$

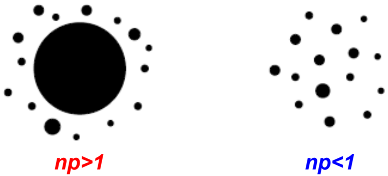
- $G_{n,p}$ is well-studied and tractable. Noteworthy properties:

P1) Degree distribution $\text{P}(d)$ is binomial with parameters $(n-1, p)$

- Large graphs have concentrated $\text{P}(d)$ with exponentially-decaying tails

P2) Phase transition on the emergence of a giant component

- If $np > 1$, $G_{n,p}$ has a giant component of size $O(n)$ w.h.p.
- If $np < 1$, $G_{n,p}$ has components of size only $O(\log n)$ w.h.p.



P3) Small clustering coefficient $O(n^{-1})$ and short diameter $O(\log n)$ w.h.p.

Generalized random graph models

- Recipe for generalization of Erdös-Renyi models

  $\Rightarrow$ Specify $\mathcal{G}$ of fixed order N_v, possessing a desired characteristic

  $\Rightarrow$ Assign equal probability to each graph $G \in \mathcal{G}$

- Configuration model: fixed degree sequence $\{d_{(1)}, \ldots, d_{(N_v)}\}$
  - Size fixed under this model, since $N_e = \bar{d} N_v / 2 \Rightarrow \mathcal{G} \subset \mathcal{G}_{N_v, N_e}$
  - Equivalent to specifying model via conditional distribution on $\mathcal{G}_{N_v, N_e}$

- Configuration models useful as reference, i.e., 'null' models

  Ex: compare observed G with $G' \in \mathcal{G}$ having power law $\mathrm{P}(d)$

  Ex: expected group-wise edge counts in modularity measure

Results on the configuration model

P1) Phase transition on the emergence of a giant component

- Condition depends on first two moments of given $\mathrm{P}(d)$
- Giant component has size $O(N_v)$ as in $G_{N_v,p}$

M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random Struct. and Alg.*, vol. 6, pp. 161-180, 1995

P2) Clustering coefficient vanishes slower than in $G_{N_v,p}$

M. Newman et al, "Random graphs with arbitrary degree distrbutions and their applications", *Physical Rev. E*, vol. 64, p. 26,118, 2001

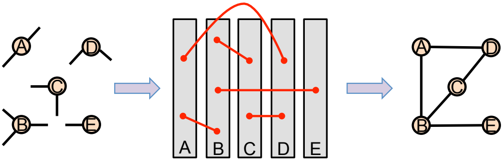
P3) Special case of given power-law degree distribution $\mathrm{P}(d) \sim Cd^{-\alpha}$

- For $\alpha \in (2,3)$, short diameter $O(\log N_v)$ as in $G_{N_v,p}$

F. Chung and L. Lu, "The average distances in random graphs with given expected degrees," *PNAS*, vol. 99, pp. 15,879-15,882, 2002

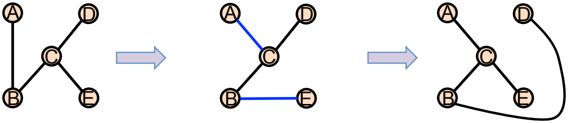
Simulating generalized random graphs

- Matching algorithm



Given: nodes with spokes   Randomly match mini-nodes   Sample graph

- Switching algorithm



Initialize   Randomly switch a pair of edges
Repeat ~$100N_e$ times   Sample graph

Task 1: Model-based estimation in network graphs

- Consider a sample G^* of a population graph $G(V, E)$
  - $\Rightarrow$ Suppose a given characteristic $\eta(G)$ is of interest
  - $\Rightarrow$ Q: Useful estimate $\hat{\eta} = \hat{\eta}(G^*)$ of $\eta(G)$?
- Statistical inference in sampling theory via design-based methods
  - $\Rightarrow$ Only source of randomness is due to the sampling design
- Augment this perspective to include a model-based component
  - Assume G drawn uniformly from the collection $\mathcal{G}$, prior to sampling
- Inference on $\eta(G)$ should incorporate both randomness due to
  - $\Rightarrow$ Selection of G from $\mathcal{G}$ and sampling G^* from G

Example: size of a "hidden population"

- Directed graph $G(V, E)$, V the members of the hidden population
  - $\Rightarrow$ Graph describing willingness to identify other members
  - $\Rightarrow$ Arc (i, j) when ask individual i, mentions j as a member
- For given V, model G as drawn from a collection $\mathcal{G}$ of random graphs
  - $\Rightarrow$ Independently add arcs between vertex pairs w.p. $p_{\mathcal{G}}$
- Graph G^* obtained via one-wave snowball sampling, i.e., $V^* = V_0^* \cup V_1^*$
  - $\Rightarrow$ Initial sample V_0^* obtained via BS from V with probability p_0
- Consider the following RVs of interest
  - $N = |V_0^*|$: size of the initial sample
  - M_1: number of arcs among individuals in V_0^*
  - M_2: number of arcs from individuals in V_0^* to individuals in V_1^*
- Snowball sampling yields measurements n, m_1, and m_2 of these RVs

Method of moments estimator

- Method of moments: now $A_{ij} = \mathbb{I}\{(i,j) \in E\}$ also a RV

$$\mathbb{E}[N] = \mathbb{E}\left[\sum_i \mathbb{I}\{i \in V_0^*\}\right] = N_v p_0 = n$$

$$\mathbb{E}[M_1] = \mathbb{E}\left[\sum_j \sum_{i \neq j} \mathbb{I}\{i \in V_0^*\}\mathbb{I}\{j \in V_0^*\}A_{ij}\right] = N_v(N_v - 1)p_0^2 p_{\mathcal{G}} = m_1$$

$$\mathbb{E}[M_2] = \mathbb{E}\left[\sum_j \sum_{i \neq j} \mathbb{I}\{i \in V_0^*\}\mathbb{I}\{j \notin V_0^*\}A_{ij}\right] = N_v(N_v - 1)p_0(1 - p_0)p_{\mathcal{G}} = m_2$$

- Expectation w.r.t. randomness in selecting G and sample V_0^*. Solution:

$$\hat{p}_0 = \frac{m_1}{m_1 + m_2},\ \hat{p}_{\mathcal{G}} = \frac{m_1(m_1 + m_2)}{n[(n-1)m_1 + nm_2]},\ \text{and}\ \hat{N}_v = n\left(\frac{m_1 + m_2}{m_1}\right)$$

$\Rightarrow$ Same estimates for p_0 and N_v as in the design-based approach

Directly modeling $\eta(G)$

- So far considered modeling G for model-based estimation of $\eta(G)$

  $\Rightarrow$ Classical random graphs typical in social networks research

- Alternatively, one may specify a model for $\eta(G)$ directly

Example

- Estimate the power-law exponent $\eta(G) = \alpha$ from degree counts
- A power law implies the linear model $\log \mathrm{P}(d) = C - \alpha \log d + \epsilon$

  $\Rightarrow$ Could use a model-based estimator such as least squares

- Better form the MLE for the model $f(d;\alpha) = \frac{\alpha-1}{d_{\min}}\left(\frac{d}{d_{\min}}\right)^{-\alpha}$

$$\text{Hill estimator} \quad \Rightarrow \hat{\alpha} = 1 + \left[\frac{1}{N_v}\sum_{i=1}^{N_v} \log\left(\frac{d_i}{d_{\min}}\right)\right]^{-1}$$

Task 2: Assessing significance in network graphs

- Consider a graph G^{obs} derived from observations
- Q: Is a structural characteristic $\eta(G^{obs})$ significant, i.e., unusual?
  - $\Rightarrow$ Assessing significance requires a frame of reference, or null model
  - $\Rightarrow$ Random graph models often used in setting up such comparisons
- Define collection $\mathcal{G}$, and compare $\eta(G^{obs})$ with values $\{\eta(G) : G \in \mathcal{G}\}$
  - $\Rightarrow$ Formally, construct the reference distribution

$$\mathrm{P}_{\eta,\mathcal{G}}(t) = \frac{|\{G \in \mathcal{G} : \eta(G) \leq t\}|}{|\mathcal{G}|}$$

- If $\eta(G^{obs})$ found to be sufficiently unlikely under $\mathrm{P}_{\eta,\mathcal{G}}(t)$
  - $\Rightarrow$ Evidence against the null H_0: G^{obs} is a uniform draw from $\mathcal{G}$

Example: Zachary's karate club

- Zachary's karate club has clustering coefficient $\text{cl}(G^{obs}) = 0.2257$

  $\Rightarrow$ Random graph models to assess whether the value is unusual

- Construct two 'comparable' abstract frames of reference
  1) Collection $\mathcal{G}_1$ of random graphs with same $N_v = 34$ and $N_e = 78$
  2) Add the constraint that $\mathcal{G}_2$ has the same degree distribution as G^{obs}

- $|\mathcal{G}_1| \approx 8.4 \times 10^{96}$ and $|\mathcal{G}_2|$ much smaller, but still large

  $\Rightarrow$ Enumerating $\mathcal{G}_1$ intractable to obtain $\text{P}_{\eta,\mathcal{G}_1}(t)$ exactly

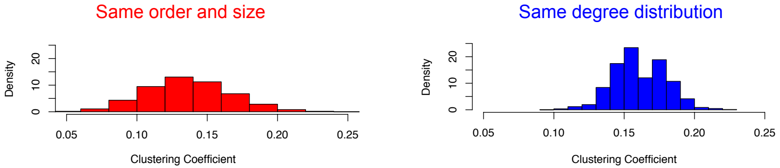
- Instead use simulations to approximate both distributions

  $\Rightarrow$ Draw 10,000 uniform samples G from each $\mathcal{G}_1$ and $\mathcal{G}_2$

  $\Rightarrow$ Calculate $\eta(G) = \text{cl}(G)$ for each sample, plot histograms

Example: Zachary's karate club (cont.)

- Plot histograms to approximate the distributions

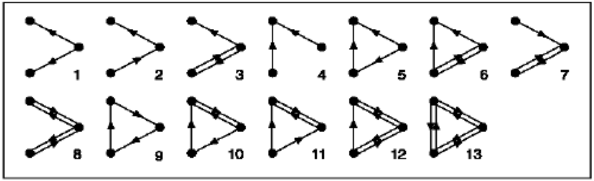


- Unlikely to see a value $\mathrm{cl}(G^{obs}) = 0.2257$ under both graph models

  Ex: only 3 out of 10,000 samples from $\mathcal{G}_1$ had $\mathrm{cl}(G) > 0.2257$

- Strong evidence to reject G^{obs} obtained as sample from $\mathcal{G}_1$ or $\mathcal{G}_2$

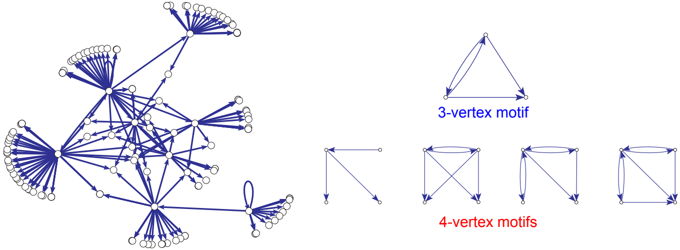
- Related use of random graph models is for detecting network motifs
  - ⇒ Find the simple 'building blocks' of a large complex network
- **Def:** Network motifs are small subgraphs occurring far more frequently in a given network than in comparable random graphs
- Ex: there are $L_3 = 13$ different connected 3-vertex subdigraphs



- Let N_i be the count in G of the i-th type k-vertex subgraph, $i = 1, \ldots, L_k$
  - ⇒ Each value N_i can be compared to a suitable reference $\mathbb{P}_{N_i,\mathcal{G}}$
  - ⇒ Subgraphs for which N_i is extreme are declared as network motifs

Example: AIDS blog network

- AIDS blog network G^{obs} with $N_v = 146$ bloggers and $N_e = 183$ links
  - $\Rightarrow$ Examined evidence for motifs of size $k = 3$ and 4 vertices



- Simulated 10,000 digraphs using a switching algorithm
  - $\Rightarrow$ Fixed in- and out-degree sequences, mutual edges as in G^{obs}
  - $\Rightarrow$ Constructed approximate reference distributions $\mathrm{P}_{N_i,\mathcal{G}}(t)$
- Ex: two bloggers with a mutual edge and a common 'authority'

Challenges in detecting motifs

- Individual motifs frequently overlap with other copies of itself

  $\Rightarrow$ May require them to be frequent and mostly disjoint subgraphs

- With large graphs come significant computational challenges

  $\Rightarrow$ Number of different potential motifs L_k grows fast with k

  Ex: Connected subdigraphs $L_3 = 13$, $L_4 = 199$, $L_5 = 9364$

- May sample subgraphs H along with the HT estimation framework

$$\hat{N}_i = \sum_{H \text{ of type } i} \pi_H^{-1}$$

Small-world models

- Random graph models
- Small-world models
- Network-growth models
- Exponential random graph models
- Latent network models
- Random dot product graphs

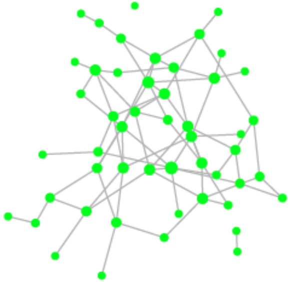
Models for real-world networks

- Arguably the most important innovation in modern graph modeling

Traditional random graph models

Transition →

Models mimicking observed ``real-world'' properties



A "small" world?

- Six degrees of separation popularized by a play [Guare'90]
  - ⇒ Short paths between us and everyone else on the planet
  - ⇒ Term relatively new, the concept has a long history
- Traced back to F. Karinthy in the 1920s
  - ⇒ 'Shrinking' modern world due to increased human connectedness
  - ⇒ Challenge: find someone whose distance from you is > 5
  - ⇒ Inspired by G. Marconi's Nobel prize speech in 1909
- First mathematical treatment [Kochen-Pool'50]
  - ⇒ Formally modeled the mechanics of social networks
  - ⇒ But left 'degrees of separation' question unanswered
- Chain of events led to a groundbreaking experiment [Milgram'67]

Milgram's experiment

- Q1: What is the typical geodesic distance between two people?

  ⇒ Experiment on the global friendship (social) network

  ⇒ Cannot measure in full, so need to probe explicitly

- S. Milgram's ingenious small-world experiment in 1967
  - 296 letters sent to people in Wichita, KS and Omaha, NE
  - Letters indicated a (unique) contact person in Boston, MA
  - Asked them to forward the letter to the contact, following rules

- **Def:** friend is someone known on a first-name basis

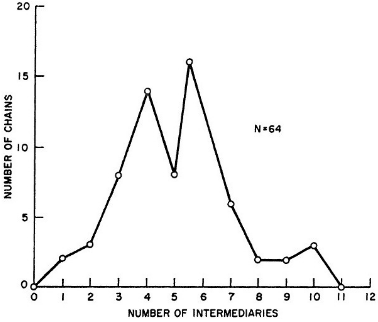
  Rule 1: If contact is a friend then send her the letter; else

  Rule 2: Relay to friend most-likely to be a contact's friend

- Q2: How many letters arrived? How long did they take?

Milgram's experimental results

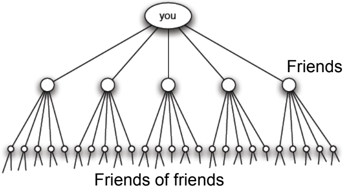
- 64 of 296 letter reached the destination, average path length $\bar{\ell} = 6.2$
  - $\Rightarrow$ Inspiring Guare's '6 degrees of separation'
- Conclusion: short paths connect arbitrary pairs of people

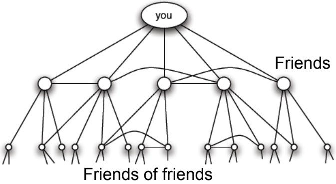


- S. Milgram, "The small-world problem," *Psychology Today*, vol. 2, pp. 60-67, 1967

Moment to reflect

- Milgram demonstrated that short paths are in abundance
- Q: Is the small-world theory reasonable? Sure, e.g., assumes:
  - We have 100 friends, each of them has 100 other friends, . . .
  - After 5 degrees we get 10^{10} friends $>$ twice the Earth's population





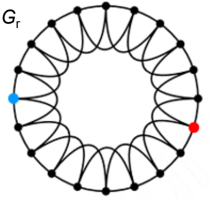
- Not a realistic model of social networks exhibiting:

  $\Rightarrow$ Homophily [Lazarzfeld'54]

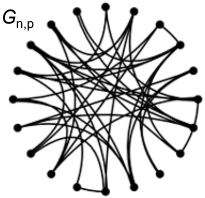
  $\Rightarrow$ Triadic closure [Rapoport'53]

- Q: How can networks be highly-structured locally **and** globally small?

Structure and randomness as extremes



High clustering and diameter



Low clustering and diameter

- One-dimensional regular lattice G_r on N_v vertices
  - Each node is connected to its $2r$ closest neighbors (r to each side)

  Structure yields high clustering and high diameter

$$\text{cl}(G_r) = \frac{3r-3}{4r-2} \text{ and } \text{diam}(G_r) = \frac{N_v}{2r}$$

- Other extreme is a $G_{N_v,p}$ random graph with $p = O(N_v^{-1})$

  Randomness yields low clustering and low diameter

$$\text{cl}(G_{N_v,p}) = O(N_v^{-1}) \text{ and } \text{diam}(G_{N_v,p}) = O(\log N_v)$$

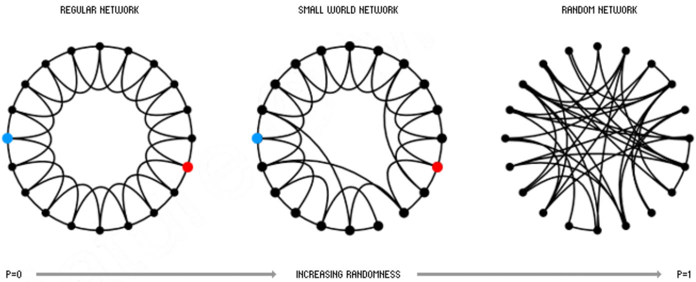
The Watts-Strogatz model

- Small-world model: blend of structure with little randomness

  S1: Start with regular lattice that has desired clustering

  S2: Introduce randomness to generate shortcuts in the graph

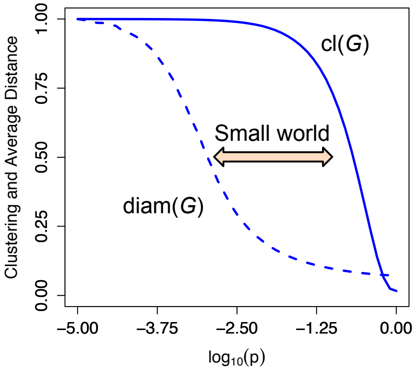
  ⇒ Each edge is randomly rewired with (small) probability p



- Rewiring interpolates between the regular and random extremes

Numerical results

- Simulate Watts-Strogatz model with $N_v = 1,000$ and $r = 6$
  - Rewiring probability p varied from 0 (lattice G_r) to 1 (random $G_{N_v,p}$)
  - Normalized $\text{cl}(G)$ and $\text{diam}(G)$ to maximum values ($p = 0$)



- Broad range of $p \in [10^{-3}, 10^{-1}]$ yields small $\text{diam}(G)$ and high $\text{cl}(G)$

Closing remarks

- Structural properties of Watts-Strogatz model [Barrat-Weigt'00]

  P1: Large N_v analysis of clustering coefficient

$$\mathrm{cl}(G) \approx \frac{3r-3}{4r-2}(1-p^3) = \mathrm{cl}(G_r)(1-p^3)$$

  P2: Degree distribution concentrated around $2r$

- Small-world graph models of interest across disciplines
- Particularly relevant to 'communication' in a broad sense

  $\Rightarrow$ Spread of news, gossip, rumors

  $\Rightarrow$ Spread of natural diseases and epidemics

  $\Rightarrow$ Search of content in peer-to-peer networks

Network-growth models

Random graph models

Small-world models

Network-growth models

Exponential random graph models

Latent network models

Random dot product graphs

Time-evolving networks

- Many networks grow or otherwise evolve in time

  Ex: Web, scientific citations, Twitter, genome ...

- General approach to model construction mimicking network growth
  - Specify simple mechanisms for network dynamics
  - Study emergent structural characteristics as time $t \to \infty$

- Q: Do these properties match observed ones in real-world networks?

- Two fundamental and popular classes of growth processes

  $\Rightarrow$ Preferential attachment models

  $\Rightarrow$ Copying models

- Tenable mechanisms for popularity and gene duplication, respectively

Preferential attachment model

- Simple model for the creation of e.g., links among Web pages
- Vertices are created one at a time, denoted $1, \ldots, N_v$
- When node j is created, it makes a single arc to i, $1 \leq i < j$
- Creation of (j, i) governed by a probabilistic rule:
  - With probability p, j links to i chosen uniformly at random
  - With probability $1 - p$, j links to i with probability $\propto d_i^{in}$
- The resulting graph is directed, each vertex has $d_v^{out} = 1$
- Preferential attachment model leads to "rich-gets-richer" dynamics

  $\Rightarrow$ Arcs formed preferentially to (currently) most popular nodes

  $\Rightarrow$ Prob. that i increases its popularity $\propto$ i's current popularity

Preferential attachment yields power laws

Theorem

The preferential attachment model gives rise to a power-law in-degree distribution with exponent $\alpha = 1 + \frac{1}{1-p}$*, i.e.,*

$$\mathrm{P}\left(d^{in} = d\right) \propto d^{-\left(1+\frac{1}{1-p}\right)}$$

- Key: "j links to i with probability $\propto d_i^{in}$" equivalent to copying, i.e., "j chooses k uniformly at random, and links to i if $(k, i) \in E$"
- Reflect: Copy other's decision vs. independent decisions in $G_{n,p}$
- As $p \to 0$ $\Rightarrow$ Copying more frequent $\Rightarrow$ Smaller $\alpha \to 2$
  - Intuitive: more likely to see extremely popular pages (heavier tail)

The Barabási-Albert model

- Barabási-Albert (BA) model is for undirected graphs
- Initial graph $G_{BA}(0)$ of $N_v(0)$ vertices and $N_e(0)$ edges $(t = 0)$
- For $t = 1, 2, \ldots$ current graph $G_{BA}(t-1)$ grows to $G_{BA}(t)$ by:
  - Adding a new vertex u of degree $d_u(t) = m \geq 1$
  - The m new edges are incident to m different vertices in $G_{BA}(t-1)$
  - New vertex u is connected to $v \in V(t-1)$ w.p.

$$\mathrm{P}\left((u,v) \in E(t)\right) = \frac{d_v(t-1)}{\sum_{v'} d_{v'}(t-1)}$$

- Vertices connected to u preferentially towards higher degrees

  $\Rightarrow G_{BA}(t)$ has $N_v(t) = N_v(0) + t$ and $N_e(t) = N_e(0) + tm$

- A. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509-512, 1999

Linearized chord diagram

- BA model ambiguous in how to select m vertices $\propto$ to their degree

  $\Rightarrow$ Joint distribution not specified by marginal on each vertex

- Linearzied chord diagram (LCD) model removes ambiguities

- For $m = 1$, start with $G_{LCD}(0)$ consisting of a vertex with a self-loop
- For $t = 1, 2, \ldots$ current graph $G_{LCD}(t-1)$ grows to $G_{LCD}(t)$ by:
  - Adding a new vertex v_t with an edge to $v_s \in V(t)$
  - Vertex v_s, $1 \leq s \leq t$ is chosen w.p.

$$\mathrm{P}\left(s=j\right) = \begin{cases} \frac{d_{v_j}(t-1)}{2t-1}, & \text{if } 1 \leq j \leq t-1, \\ \frac{1}{2t-1}, & \text{if } j = t \end{cases}$$

- For $m > 1$ simply run the above process m times for each t
  - Collapse all created vertices into a single one, retaining edges

- A. Bollobás et al, "The degree sequence of a scale-free random graph process," *Random Struct. and Alg.*, vol. 18, pp. 279-290, 2001

Properties of the LCD model

P1) The LCD model allows for loops and multi-edges, occurring rarely

P2) $G_{LCD}(t)$ has power-law degree distribution with $\alpha = 3$, as $t \to \infty$

P3) The BA model yields connected graphs if $G_{BA}(0)$ connected

$\Rightarrow$ Not true for the LCD model, but $G_{LCD}(t)$ connected w.h.p.

P4) Small-world behavior

$$\text{diam}(G_{LCD}(t)) = \begin{cases} O(\log N_v(t)), & m = 1 \\ O\left(\frac{\log N_v(t)}{\log \log N_v(t)}\right), & m > 1 \end{cases}$$

P5) Unsatisfactory clustering, since small for $m > 1$

$$\mathbb{E}\left[\text{cl}(G_{LCD}(t))\right] \approx \frac{m-1}{8} \frac{(\log N_v(t))^2}{N_v(t)}$$

$\Rightarrow$ Marginally better than $O(N_v^{-1})$ in classical random graphs

Copying models

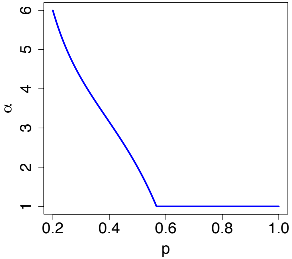
- Copying is another mechanism of fundamental interest

  Ex: gene duplication to re-use information in organism's evolution
- Different from preferential attachment, but still results in power laws
- Initialize with a graph $G_C(0)$ $(t = 0)$
- For $t = 1, 2, \ldots$ current graph $G_C(t-1)$ grows to $G_C(t)$ by:
  - Adding a new vertex u
  - Choosing vertex $v \in V(t-1)$ with uniform probability $\frac{1}{N_v(t-1)}$
  - Joining vertex u with v's neighbors independently w.p. p
- Case $p = 1$ leads to full duplication of edges from an existing node
- F. Chung et al, "Duplication models for biological networks," *Journal of Computational Biology*, vol. 10, pp. 677-687, 2003

Asymptotic degree distribution

- Degree distribution tends to a power law w.h.p. [Chung et al'03]
  - $\Rightarrow$ Exponent α is the plotted solution to the equation

$$p(\alpha - 1) = 1 - p^{\alpha - 1}$$



- Full duplication does not lead to power-law behavior; but does if
  - $\Rightarrow$ Partial duplication performed a fraction $q \in (0, 1)$ of times

Fitting network growth models

- Most common practical usage of network growth models is predictive
  Goal: compare characteristics of G^{obs} and $G(t)$ from the models
- Little attempt to date to fit network growth models to data
  $\Rightarrow$ Expected due to simplicity of such models
  $\Rightarrow$ Still useful to estimate e.g., the duplication probability p
- To fit a model ideally would like to observe a sequence $\{G^{obs}(\tau)\}_{\tau=1}^{t}$
  $\Rightarrow$ Unfortunately, such dynamic network data is still fairly elusive
- Q: Can we fit a network growth model to a single snap-shot G^{obs}?
- A: Yes, if we leverage the Markovianity of the growth process

Duplication-attachment models

- Similar to all network growth models described so far, suppose:

  As1: A single vertex is added to $G(t-1)$ to create $G(t)$; and
  As2: The manner in which it is added depends only on $G(t-1)$

- In other words, we assume $\{G(t)\}_{t=0}^{\infty}$ is a Markov chain

- Let graph $\delta(G(t), v)$ be obtained by deleting v and its edges from $G(t)$

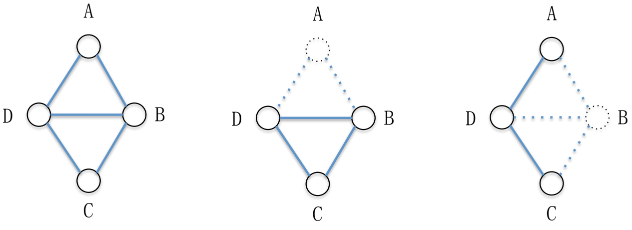
- **Def:** vertex v is removable if $G(t)$ can be obtained from $\delta(G(t), v)$ via copying. If $G(t)$ has no removable vertices, we call it irreducible

- The class of duplication-attachment (DA) models satisfies:

  (i) The initial graph $G(0)$ is irreducible; and
  (ii) $\mathrm{P}_\theta(G(t) \,|\, G(t-1)) > 0 \Leftrightarrow G(t)$ obtained by copying a vertex in $G(t-1)$

- C. Wiuf et al, "A likelihood approach to analysis of network data," *PNAS*, vol. 103, pp. 7566-7570, 2006

Example: reducible graph



$G(t)$ $\delta(G(t), v_A)$ $\delta(G(t), v_B)$

- Vertex v_A is removable (likewise v_c by symmetry)
  - $\Rightarrow$ Obtain $G(t)$ from $\delta(G(t, v_a))$ by copying v_c
- This implies that $G(t)$ is reducible
  - $\Rightarrow$ Notice though that v_B or v_D are not removable

MLE for DA model parameters

- Suppose that $G^{obs} = G(t)$ represents the observed network graph
- The likelihood for the parameter θ is recursively given by

$$\mathcal{L}(\theta; G(t)) = \frac{1}{t} \sum_{v \in \mathcal{R}_{G(t)}} \mathrm{P}_\theta \left(G(t) \,\middle|\, \delta(G(t), v) \right) \mathcal{L}(\theta; \delta(G(t), v))$$

$\Rightarrow$ $\mathcal{R}_{G(t)}$ is the set of all removable nodes in $G(t)$

- The MLE for θ is thus defined as

$$\hat{\theta} = \arg\max_\theta \mathcal{L}(\theta; G(t))$$

$\Rightarrow$ Computing $\mathcal{L}(\theta; G(t))$ non-trivial, even for modest-size graphs

- Monte Carlo methods to approximate $\mathcal{L}(\theta; G(t))$ [Wiupf et al'06]

  $\Rightarrow$ Open issues: vector $\boldsymbol{\theta}$, other growth models, scalability

Exponential random graph models

Random graph models

Small-world models

Network-growth models

Exponential random graph models

Latent network models

Random dot product graphs

Statistical network graph models

- Good statistical network graph models should be [Robbins-Morris'07]:
  - ⇒ Estimable from and reasonably representative of the data
  - ⇒ Theoretically plausible about the underlying network effects
  - ⇒ Discriminative among competing effects to best explain the data
- Network-based versions of canonical statistical models
  - ⇒ Regression models - Exponential random graph models (ERGMs)
  - ⇒ Latent variable models - Latent network models
  - ⇒ Mixture models - Stochastic block models
- Focus here on ERGMs, also known as p^* models
- G. Robbins et al., "An introduction to exponential random graph (p^*) models for social networks," *Social Networks,* vol. 29, pp. 173-191, 2007

Exponential family

- **Def:** discrete random vector $\mathbf{Z} \in \mathcal{Z}$ belongs to an exponential family if

$$\text{P}_\theta(\mathbf{Z} = \mathbf{z}) = \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{z}) - \psi(\theta)\right\}$$

  - $\boldsymbol{\theta} \in \mathbb{R}^p$ is a vector of parameters and $\mathbf{g} : \mathcal{Z} \mapsto \mathbb{R}^p$ is a function
  - $\psi(\theta)$ is a normalization term, ensuring $\sum_{\mathbf{z} \in \mathcal{Z}} \text{P}_\theta(\mathbf{z}) = 1$
  - Ex: Bernoulli, binomial, Poisson, geometric distributions

- For continuous exponential families, the pdf has an analogous form

  Ex: Gaussian, Pareto, chi-square distributions

- Exponential families share useful algebraic and geometric properties

  $\Rightarrow$ Mathematically convenient for inference and simulation

Exponential random graph model

- Let $G(V, E)$ be a random undirected graph, with $Y_{ij} := \mathbb{I}\{(i,j) \in E\}$
  - Matrix $\mathbf{Y} = [Y_{ij}]$ is the random adjacency matrix, $\mathbf{y} = [y_{ij}]$ a realization
- An ERGM specifies in exponential family form the distribution of $\mathbf{Y}$, i.e.,

$$\mathrm{P}_\theta(\mathbf{Y} = \mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right) \exp\left\{\sum_H \theta_H g_H(\mathbf{y})\right\}, \quad \text{where}$$

(i) each H is a configuration, meaning a set of possible edges in G;
(ii) $g_H(\mathbf{y})$ is the network statistic corresponding to configuration H

$$g_H(\mathbf{y}) = \prod_{y_{ij} \in H} y_{ij} = \mathbb{I}\{H \text{ occurs in } \mathbf{y}\}$$

(iii) $\theta_H \neq 0$ only if all edges in H are conditionally dependent; and
(iv) $\kappa(\boldsymbol{\theta})$ is a normalization constant ensuring $\sum_{\mathbf{y}} \mathrm{P}_\theta(\mathbf{y}) = 1$

Discussion

- Graph order N_v is fixed and given, only edges are random

  $\Rightarrow$ Assumed unweighted, undirected edges. Extensions possible

- ERGMs describe random graphs 'built-on' localized patterns
  - These configurations are the structural characteristics of interest
  - Ex: Are there reciprocity effects? Add mutual arcs as configurations
  - Ex: Are there transitivity effects? Consider triangles

- (In)dependence is conditional on all other variables (edges) in G

  $\Rightarrow$ Control configurations relevant (i.e., $\theta_H \neq 0$) to the model

- Well-specified dependence assumptions imply particular model classes

A general framework for model construction

- In positing an ERGM for a network, one implicitly follows five steps

  $\Rightarrow$ Explicit choices connecting hypothesized theory to data analysis

  Step 1: Each edge (relational tie) is regarded as a random variable

  Step 2: A dependence hypothesis is proposed

  Step 3: Dependence hypothesis implies a particular form to the model

  Step 4: Simplification of parameters through e.g., homogeneity

  Step 5: Estimate and interpret model parameters

Example: Bernoulli random graphs

- Assume edges present independently of all other edges (e.g., in $G_{n,p}$)

  $\Rightarrow$ Simplest possible (and unrealistic) dependence assumption

- For each (i,j), we assume Y_{ij} independent of Y_{uv}, for all $(u,v) \neq (i,j)$

  $\Rightarrow \theta_H = 0$ for all H involving two or more edges

- Edge configurations i.e., $g_H(\mathbf{y}) = y_{ij}$ relevant, and the ERGM becomes

$$\mathrm{P}_\theta(\mathbf{Y}=\mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right) \exp\left\{\sum_{i,j} \theta_{ij} y_{ij}\right\}$$

- Specifies that edge (i,j) present independently, with probability

$$p_{ij} = \frac{\exp(\theta_{ij})}{1+\exp(\theta_{ij})}$$

Constraints on parameters: homogeneity

- Too many parameters makes estimation infeasible from single $\mathbf{y}$

  $\Rightarrow$ Under independence have N_v^2 parameters $\{\theta_{ij}\}$. Reduction?

- Homogeneity across all G, i.e., $\theta_{ij} = \theta$ for all (i, j) yields

$$\mathrm{P}_\theta(\mathbf{Y} = \mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right) \exp\left\{\theta L(\mathbf{y})\right\}$$

  - Relevant statistic is the number of edges observed $L(\mathbf{y}) = \sum_{i,j} y_{ij}$
  - ERGM identical to $G_{n,p}$, where $p = \dfrac{\exp\theta}{1 + \exp\theta}$

Ex: suppose we know a priori that vertices fall in two sets

- Can impose homogeneity on edges within and between sets, i.e.,

$$\mathrm{P}_\theta(\mathbf{Y} = \mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right) \exp\left\{\theta_1 L_1(\mathbf{y}) + \theta_{12} L_{12}(\mathbf{y}) + \theta_2 L_2(\mathbf{y})\right\}$$

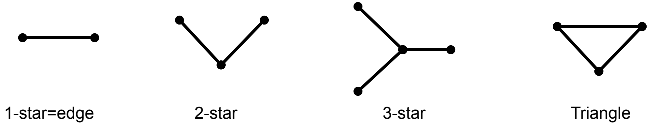
- Markov dependence notion for network graphs [Frank-Strauss'86]
  - Assumes two ties are dependent if they share a common node
  - Edge status Y_{ij} dependent on any other edge involving i or j

Theorem

Under homogeneity, G is a Markov random graph if and only if

$$P_\theta(\mathbf{Y}=\mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right)\exp\left\{\sum_{k=1}^{N_v-1}\theta_k S_k(\mathbf{y}) + \theta_\tau T(\mathbf{y})\right\}, \quad \textit{where}$$

$S_k(\mathbf{y})$ is the number of k-stars, and $T(\mathbf{y})$ the number of triangles



Alternative statistics

- Including many higher-order terms challenges estimation
  - $\Rightarrow$ High-order star effects often omitted, e.g., $\theta_k = 0$, $k \geq 4$
  - $\Rightarrow$ But these models tend to fit real data poorly. Dilemma?
- Idea: Impose parametric form $\theta_k \propto (-1)^k \lambda^{2-k}$ [Snijders et al'06]
- Combine $S_k(\mathbf{y})$, $k \geq 2$ into a single alternating k-star statistic, i.e.,

$$\text{AKS}_\lambda(\mathbf{y}) = \sum_{k=2}^{N_v-1} (-1)^k \frac{S_k(\mathbf{y})}{\lambda^{k-2}}, \quad \lambda > 1$$

- Can show $\text{AKS}_\lambda(\mathbf{y}) \propto$ the geometrically-weighted degree count

$$\text{GWD}_\gamma(\mathbf{y}) = \sum_{d=0}^{N_v-1} e^{-\gamma d} N_d(\mathbf{y}), \quad \gamma > 0$$

  - $\Rightarrow$ $N_d(\mathbf{y})$ is the number of vertices with degree d

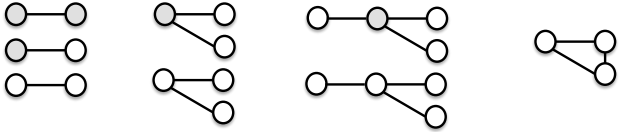
Incorporating vertex attributes

- Straightforward to incorporate vertex attributes to ERGMs
  Ex: gender, seniority in organization, protein function
- Consider a realization $\mathbf{x}$ of a random vector $\mathbf{X} \in \mathbb{R}^{N_v}$ defined on V
- Specify an exponential family form for the conditional distribution

$$\mathrm{P}_\theta\left(\mathbf{Y}=\mathbf{y} \,\middle|\, \mathbf{X}=\mathbf{x}\right)$$

$\Rightarrow$ Will include additional statistics $g(\cdot)$ of $\mathbf{y}$ and $\mathbf{x}$

- Ex: configurations for Markov, binary vertex attributes



Estimating ERGM parameters

- MLE for the parameter vector $\boldsymbol{\theta}$ in an ERGM is

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ \boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\boldsymbol{\theta}) \right\}, \quad \text{where } \psi(\boldsymbol{\theta}) := \log \kappa(\boldsymbol{\theta})$$

- Optimality condition yields

$$\mathbf{g}(\mathbf{y}) = \nabla \psi(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

- Using also that $\mathbb{E}_\theta[\mathbf{g}(\mathbf{Y})] = \nabla\psi(\boldsymbol{\theta})$, the MLE solves

$$\mathbb{E}_{\hat{\theta}}[\mathbf{g}(\mathbf{Y})] = \mathbf{g}(\mathbf{y})$$

- Unfortunately $\psi(\boldsymbol{\theta})$ cannot be computed except for small graphs

  $\Rightarrow$ Involves a summation over $2^{\binom{N_v}{2}}$ values of $\mathbf{y}$ for each $\boldsymbol{\theta}$

  $\Rightarrow$ Numerical methods needed to obtain approximate values of $\hat{\boldsymbol{\theta}}$

Proof of $\mathbb{E}[g(\mathbf{Y})] = \nabla\psi(\theta)$

- The pmf of $\mathbf{Y}$ is $\mathrm{P}_\theta(\mathbf{Y}=\mathbf{y}) = \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}$, hence

$$\begin{aligned}\mathbb{E}_\theta[g(\mathbf{Y})] &= \sum_{\mathbf{y}} g(\mathbf{y})\mathrm{P}_\theta(\mathbf{Y}=\mathbf{y})\\ &= \sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}\end{aligned}$$

- Recall $\psi(\boldsymbol{\theta}) = \log\sum_{\mathbf{y}} \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}$ and use the chain rule

$$\begin{aligned}\nabla\psi(\boldsymbol{\theta}) &= \frac{\sum_{\mathbf{y}} g(\mathbf{y})\exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}}{\sum_{\mathbf{y}} \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}} = \frac{\sum_{\mathbf{y}} g(\mathbf{y})\exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}}{\exp\psi(\theta)}\\ &= \sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}\end{aligned}$$

- The red and blue sums are identical $\Rightarrow \mathbb{E}_\theta[g(\mathbf{Y})] = \nabla\psi(\boldsymbol{\theta})$ follows

Markov chain Monte Carlo MLE

- Idea: for fixed $\boldsymbol{\theta}_0$, maximize instead the log-likelihood ratio

$$r(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta}_0) = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{y}) - [\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)]$$

- Key identity: will show that

$$\exp\{\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)\} = \mathbb{E}_{\theta_0}\left[\exp\left\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{Y})\right\}\right]$$

- Markov chain Monte Carlo MLE algorithm to search over $\boldsymbol{\theta}$

  Step 1: draw samples $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ from the ERGM under $\boldsymbol{\theta}_0$

  Step 2: approximate the above $\mathbb{E}_{\theta_0}[\cdot]$ via sample averaging

  Step 3: the logarithm of the result approximates $\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)$

  Step 4: evaluate an $\approx$ log-likelihood ratio $r(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$

- For large n, the maximum value found approximates the MLE $\hat{\boldsymbol{\theta}}$

Derivation of key identity

- Recall $\exp \psi(\boldsymbol{\theta}) = \sum_{\mathbf{y}} \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}$ to write

$$\exp\left\{\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)\right\} = \frac{\sum_{\mathbf{y}} \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}}{\exp \psi(\boldsymbol{\theta}_0)}$$

- Multiplying and dividing by $\exp\left\{\boldsymbol{\theta}_0^\top \mathbf{g}(\mathbf{y})\right\} > 0$ yields

$$\begin{aligned}
\exp\left\{\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)\right\} &= \sum_{\mathbf{y}} \exp\left\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{y})\right\} \times \frac{\exp\left\{\boldsymbol{\theta}_0^\top \mathbf{g}(\mathbf{y})\right\}}{\exp \psi(\boldsymbol{\theta}_0)} \\
&= \sum_{\mathbf{y}} \exp\left\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{y})\right\} \mathrm{P}_{\theta_0}(\mathbf{Y} = \mathbf{y}) \\
&= \mathbb{E}_{\theta_0}\left[\exp\left\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{Y})\right\}\right]
\end{aligned}$$

- Used $\exp\left\{\boldsymbol{\theta}_0^\top \mathbf{g}(\mathbf{y}) - \psi(\boldsymbol{\theta}_0)\right\}$ is the exponential family pmf $\mathrm{P}_{\theta_0}(\mathbf{Y} = \mathbf{y})$

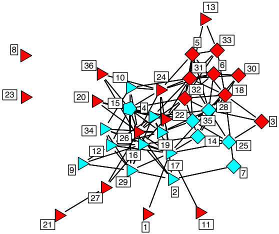
Model goodness-of-fit

- Best fit chosen from a given class of models ...
  may not be a good fit to the data if model class not rich enough
- Assessing goodness-of-fit for ERGMs

  Step 1: simulate numerous random graphs from the fitted model
  Step 2: compare high-level characteristics with those of G^{obs}
  Ex: distributions of degree, centrality, diameter
- If significant differences found in G^{obs}, conclude
  $\Rightarrow$ Systematic gap between specified model class and data
  $\Rightarrow$ Lack of goodness-of-fit
- **Take home:** model specification for ERGMs highly nontrivial
  $\Rightarrow$ Goodness-of-fit diagnostics can play key facilitating role

Lawyer collaboration network

- Network G^{obs} of working relationships among lawyers [Lazega'01]
  - Nodes are $N_v = 36$ partners, edges indicate partners worked together



- Data includes various node-level attributes:
  - Seniority (node labels indicate rank ordering)
  - Office location (triangle, square or pentagon)
  - Type of practice, i.e., litigation (red) and corporate (cyan)
  - Gender (three partners are female labeled 27, 29 and 34)
- Goal: study cooperation among social actors in an organization

Modeling lawyer collaborations

- Assess network effects $S_1(\mathbf{y}) = N_e$ and alternating k-triangles statistic

$$\text{AKT}_\lambda(\mathbf{y}) = 3T_1(\mathbf{y}) + \sum_{k=2}^{N_v-2} (-1)^{k+1} \frac{T_k(\mathbf{y})}{\lambda^{k-1}}$$

$\Rightarrow$ $T_k(\mathbf{y})$ counts sets of k individual triangles sharing a common base

- Test the following set of exogenous effects:

$$h^{(1)}(\mathbf{x}_i, \mathbf{x}_j) = \text{seniority}_i + \text{seniority}_j, \quad h^{(2)}(\mathbf{x}_i, \mathbf{x}_j) = \text{practice}_i + \text{practice}_j$$
$$h^{(3)}(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{I}\left\{\text{practice}_i = \text{practice}_j\right\}, \quad h^{(4)}(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{I}\left\{\text{gender}_i = \text{gender}_j\right\}$$
$$h^{(5)}(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{I}\left\{\text{office}_i = \text{office}_j\right\}, \quad \mathbf{h}(\mathbf{x}_i, \mathbf{x}_j) := [h^{(1)}(\mathbf{x}_i, \mathbf{x}_j), \ldots, h^{(5)}(\mathbf{x}_i, \mathbf{x}_j)]^T$$

- Resulting ERGM

$$\mathbb{P}_{\boldsymbol{\theta},\boldsymbol{\beta}}(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}) = \frac{1}{\kappa(\boldsymbol{\theta}, \boldsymbol{\beta})} \exp\left\{\theta_1 S_1(\mathbf{y}) + \theta_2 \text{AKT}_\lambda(\mathbf{y}) + \boldsymbol{\beta}^T \mathbf{g}(\mathbf{y}, \mathbf{x})\right\}$$
$$\mathbf{g}(\mathbf{y}, \mathbf{x}) = \sum_{i,j} y_{ij} \mathbf{h}(\mathbf{x}_i, \mathbf{x}_j)$$

Model fitting result

- Fitting results using the MCMC MLE approach

Parameter	Estimate	'Standard Error'
Density (θ_1)	-6.2073	0.5697
Alternating k-triangles (θ_2)	0.5909	0.0882
Seniority Main Effect (β_1)	0.0245	0.0064
Practice Main Effect (β_2)	0.3945	0.1103
Same Practice (β_3)	0.7721	0.1973
Same Gender (β_4)	0.7302	0.2495
Same Office (β_5)	1.1614	0.1952

$\Rightarrow$ Standard errors heuristically obtained via asymptotic theory

- Identified factors that may increase odds of cooperation

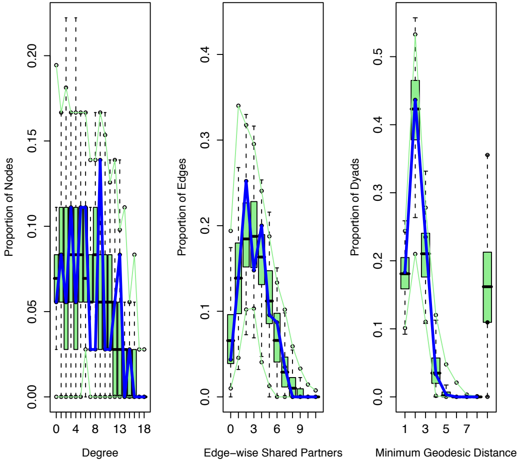
  Ex: same practice, gender and office location double odds

- Strong evidence for transitivity effects since $\hat{\theta}_2 \gg \text{se}(\hat{\theta}_2)$

  $\Rightarrow$ Something beyond basic homophily explaining such effects

Assessing goodness-of-fit

- Assess goodness-of-fit to G^{obs}
  - Sample from fitted ERGM
- Compared distributions of
  - Degree
  - Edge-wise shared partners
  - Geodesic distance
- Plots show good fit overall



Roadmap

Random graph models

Small-world models

Network-growth models

Exponential random graph models

Latent network models

Random dot product graphs

Latent network models

- Latent variables widely used to model observed data

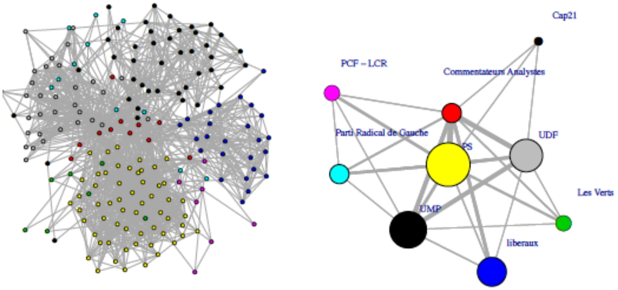
  Ex: Hidden Markov models, factor analysis

- Basic idea permeated to statistical network analysis. Two types:
  - Latent class models: unobserved class membership drives propensity towards establishing relational ties
  - Latent feature models: relational ties more likely to form among vertices that are 'closer' in some latent space

- As of now latent network models come in many flavors. Focus here:

  $\Rightarrow$ Stochastic block models (SBMs)

- French political blog network from October 2006 [Kolaczyk'17]
  - $\Rightarrow$ Consists of $N_v = 192$ blogs linked by $N_e = 1431$ edges
  - $\Rightarrow$ Colors indicate blog affiliation to a French political party



- Visual evidence of mixing of smaller subgraphs
  - $\Rightarrow$ Different rates of connections among blogs (driven by party)
  - $\Rightarrow$ Erdös-Renyi with fixed p cannot capture this structure

Stochastic block models

- SBMs explicitly parameterize the notion of communities $\mathcal{C}_1, \ldots, \mathcal{C}_Q$

  $\Rightarrow$ Connection rates π_{qr} of vertices between/within groups

SBM. Generative model for an undirected random graph $G(\mathcal{V}, \mathcal{E})$

- Fix Q. Each vertex $i \in \mathcal{V}$ independently belongs to $\mathcal{C}_q$ w.p. α_q

$$\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_Q]^\top, \quad \mathbf{1}^\top \boldsymbol{\alpha} = 1$$

- For vertices $i, j \in \mathcal{V}$, with $i \in \mathcal{C}_q$ and $j \in \mathcal{C}_r$ $\Rightarrow$ $(i, j) \in \mathcal{E}$ w.p. π_{qr}

P. W. Holland et al., "Stochastic block-models: First steps," *Social Networks*, vol. 5, pp. 109-137, 1983

Model specification and flexibility

- In other words, with $Z_{iq} = \mathbb{I}\{i \in \mathcal{C}_q\}$ and $\mathbf{Z}_i = [Z_{i1}, \ldots, Z_{iQ}]^\top$

$$\mathbf{Z}_i \overset{\text{i.i.d.}}{\sim} \text{Multinomial}(1, \boldsymbol{\alpha}),$$
$$A_{ij} \,\big|\, \mathbf{Z}_i = \mathbf{z}_i, \mathbf{Z}_j = \mathbf{z}_j \sim \text{Bernoulli}(\pi_{\mathbf{z}_i, \mathbf{z}_j})$$

  for $1 \leq i, j \leq N_v$, where $A_{ij} = A_{ji}$ and $A_{ii} \equiv 0$

- Parameters: Q mixing weights α_q and $Q(Q+1)/2$ connection probs. π_{qr}
- Mixture of classical random graph models

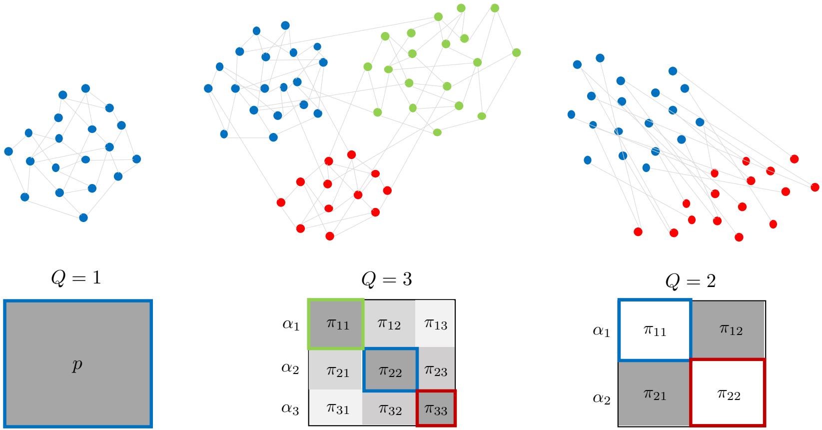
$$\text{P}\left(A_{ij} = 1\right) = \sum_{1 \leq q, r \leq Q} \alpha_q \alpha_r \pi_{qr}$$

  $\Rightarrow$ More flexible to capture the structure of observed networks

  $\Rightarrow$ May face issues of identifiability [Allman et al'11]

- Emergence of giant component, size distribution of groups [Söderberg'03]

Model specification and flexibility (cont.)



- Mixtures of Erdös-Renyi models can be surprisingly flexible

Plausibility

- Good statistical network graph models should be [Robbins-Morris'07]:
  - ⇒ Estimable from and reasonably representative of the data
  - ⇒ Theoretically plausible about the underlying network effects
- **Q:** How appropriate are latent network models? Are they plausible?
- **Q:** Can we approximate well an observed graph G^{obs} with an SBM?
  - ⇒ A variant of the Szemerédi regularity lemma useful here

C. Borgs et al, "Graph limits and parameter testing," *Symposium on Theory of Computing*, 2006

Cut distance

- Discussing approximation notions requires a distance between graphs
- **Def:** For graphs $G(\mathcal{V}, \mathcal{E})$ and $G'(\mathcal{V}', \mathcal{E}')$ with $|\mathcal{V}| = |\mathcal{V}'| = N_v$, the cut distance is given by

$$d_{\square}(G, G') = \frac{1}{N_v^2} \max_{\mathcal{S}, \mathcal{T} \in \{1, \ldots, N_v\}} \left| \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{T}} (A_{ij} - A'_{ij}) \right|$$

$\Rightarrow$ One can show the quantity $d_{\square}(\cdot, \cdot)$ is a formal metric

- Defining and studying properties of graph distances is a timely topic

B. Bollobás and O. Riordan, "Sparse graphs: Metrics and random models," *Random Structures & Algorithms*, vol. 39, 2011

An approximation result

- Let $\mathcal{P} = \{\mathcal{V}_1, \ldots, \mathcal{V}_Q\}$ partition the vertices $\mathcal{V}$ of G into Q classes
- Define the complete graph G_P with vertex set $\mathcal{V}$ and edge weights

$$w_{ij}(G_P) = \frac{1}{|\mathcal{V}_q||\mathcal{V}_r|} \sum_{u \in \mathcal{V}_q} \sum_{v \in \mathcal{V}_r} A_{uv}, \quad i \in \mathcal{V}_q, j \in \mathcal{V}_r$$

$\Rightarrow$ Expectation of a $Q-$class block model approximation to G

$\Rightarrow$ Probability an edge joins i, j is just $w_{ij}(G_P)$

Theorem: For every $\epsilon > 0$ and every graph $G(\mathcal{V}, \mathcal{E})$, there exists a partition $\mathcal{P}$ of $\mathcal{V}$ into $Q \leq 2^{\frac{2}{\epsilon^2}}$ classes such that $d_{\square}(G, G_P) \leq \epsilon$.

- Justifies the claim that an SBM can approximate well an arbitrary graph

$\Rightarrow$ The upper bound on Q can be prohibitively large

Estimating SBM parameters

- SBMs defined up to parameters $\{\alpha_q\}_{q=1}^{Q}$ and $\{\pi_{qr}\}_{1\leq q,r\leq Q}$
- Conceptually useful to think about two sets of 'observations'
  - ⇒ Latent class labels: $\mathbf{Z} = \{\{Z_{iq}\}_{q=1}^{Q}\}_{i\in\mathcal{V}}$, where $Z_{iq} = \mathbb{I}\{i \in \mathcal{C}_q\}$
  - ⇒ Relational ties: $\mathbf{A} = [A_{ij}]$, where $A_{ij} = \mathbb{I}\{(i,j) \in \mathcal{E}\}$
- **But we only observe A**, recall $\mathbf{Z}$ are latent. Q assumed given
  - ⇒ Interest both in parameter estimation and in vertex clustering

Model-based community detection

Suppose G adheres to an SBM with Q classes. Predict class membership labels $\mathbf{Z} = \{\{Z_{iq}\}_{q=1}^{Q}\}_{i\in\mathcal{V}}$, given observations $\mathbf{A} = \mathbf{a}$.

Maximum likelihood estimation

- If we were to observe $\mathbf{A} = \mathbf{a}$ and $\mathbf{Z} = \mathbf{z}$, the log-likelihood would be

$$\ell(\mathbf{a}, \mathbf{z}; \boldsymbol{\theta}) = \sum_i \sum_q z_{iq} \log \alpha_q + \frac{1}{2} \sum_{i \neq j} \sum_{q,r} z_{iq} z_{jr} \log b(A_{ij}; \pi_{qr})$$

$\Rightarrow$ Defined $\boldsymbol{\theta} = \{\{\alpha_q\}, \{\pi_{qr}\}\}$ and $b(a; \pi) = \pi^a (1 - \pi)^{1-a}$

- **But we do not.** Instead have to work with the observed data likelihood

$$\ell(\mathbf{a}; \boldsymbol{\theta}) = \log \left(\sum_{\mathbf{z}} \exp \{\ell(\mathbf{a}, \mathbf{z}; \boldsymbol{\theta})\} \right)$$

$\Rightarrow$ Unfortunately, evaluation of $\ell(\mathbf{a}; \boldsymbol{\theta})$ is typically intractable

- Mixture model viewpoint suggests an E-M procedure [Snijders'97]

$\Rightarrow$ Alternate between estimation of $\mathbb{E}\left[Z_{iq} \,\middle|\, \mathbf{A} = \mathbf{a}\right]$ and $\boldsymbol{\theta}$

$\Rightarrow$ Does not scale beyond $Q = 2$, $\mathrm{P}\left(\mathbf{Z} \,\middle|\, \mathbf{A} = \mathbf{a}\right)$ expensive

Variational maximum likelihood

- Variational approach to optimize a lower bound of $\ell(\mathbf{a}; \boldsymbol{\theta})$, namely

$$J(R_{\mathbf{a}}; \boldsymbol{\theta}) = \ell(\mathbf{a}; \boldsymbol{\theta}) - \mathrm{KL}(R_{\mathbf{a}}(\mathbf{Z}), \mathrm{P}\left(\mathbf{Z} \,\middle|\, \mathbf{A} = \mathbf{a}\right))$$

  - KL denotes de Kullback–Leibler divergence
  - $R_{\mathbf{a}}(\mathbf{Z})$ is a tractable approximation of $\mathrm{P}\left(\mathbf{Z} \,\middle|\, \mathbf{A} = \mathbf{a}\right)$

- Mean field approximation to the conditional distribution

$$R_{\mathbf{a}}(\mathbf{Z}) = \prod_{i=1}^{N_v} h(\mathbf{Z}_i; \boldsymbol{\tau}_i)$$

  - $h(\cdot; \boldsymbol{\tau}_i)$: multinomial pmf with parameter $\boldsymbol{\tau}_i = [\tau_{i1}, \ldots, \tau_{iQ}]^\top$

J. J. Daudin et al, "A mixture model for random graphs," *Stat. Comput.*, vol. 18, 2008

Alternating maximization algorithm

Proposition: Given $\boldsymbol{\theta}$, the optimal variational parameters $\{\hat{\boldsymbol{\tau}}_i\} = \text{argmax}_{\{\boldsymbol{\tau}_i\}} J(R_{\mathbf{a}}; \{\boldsymbol{\tau}_i\}, \boldsymbol{\theta})$ satisfy the following fixed-point relation

$$\hat{\tau}_{iq} \propto \alpha_q \prod_{j \neq i} \prod_r b(A_{ij}; \pi_{qr})^{\hat{\tau}_{jr}}$$

Given $\{\boldsymbol{\tau}_i\}$, the values of $\boldsymbol{\theta}$ that maximize $J(R_{\mathbf{a}}; \{\boldsymbol{\tau}_i\}, \boldsymbol{\theta})$ are

$$\hat{\alpha}_q = \frac{1}{N_v} \sum_i \tau_{iq}, \quad \hat{\pi}_{qr} = \sum_{i \neq j} \tau_{iq}\tau_{jr} A_{ij} \Big/ \sum_{i \neq j} \tau_{iq}\tau_{jr}$$

- Algorithm alternates between updates of $\boldsymbol{\theta}$ and $\{\boldsymbol{\tau}_i\}$ as follows

$$\boldsymbol{\theta}[k+1] = \underset{\boldsymbol{\theta}}{\text{argmax}}\, J(R_{\mathbf{a}}; \{\boldsymbol{\tau}_i[k]\}, \boldsymbol{\theta})$$

$$\{\boldsymbol{\tau}_i[k+1]\} = \underset{\{\boldsymbol{\tau}_i\}}{\text{argmax}}\, J(R_{\mathbf{a}}; \{\boldsymbol{\tau}_i\}, \boldsymbol{\theta}[k+1])$$

  - The sequence of J values is non-decreasing [Daudin et al'08]
  - Consistency results available as $N_v \to \infty$, Q fixed [Celisse et al'12]

Choice of the number of classes

- Number of classes Q often unknown and should be estimated
  - $\Rightarrow$ Use principles of Bayesian model selection
  - $\Rightarrow$ Prior $g(\boldsymbol{\theta} \,|\, m_Q)$ on $\boldsymbol{\theta}$ given the SBM m_Q has Q classess
- Integrated Classification Likelihood (ICL) criterion yields

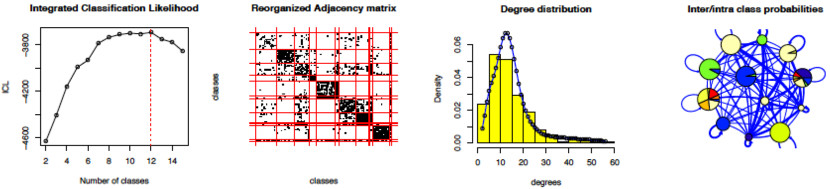
$$\begin{aligned} \text{ICL}(m_Q) = \; & \max_{\boldsymbol{\theta}} \log \mathcal{L}(\mathbf{a}, \hat{\mathbf{z}}(\boldsymbol{\theta}) \,|\, \boldsymbol{\theta}, m_Q) \\ & - \frac{Q(Q+1)}{4} \log \frac{N_v(N_v - 1)}{2} - \frac{Q-1}{2} \log N_v \end{aligned}$$

- Asymptotic approximation of the complete-data integrated likelihood

$$\mathcal{L}(\mathbf{a}, \mathbf{z} \,|\, m_Q) = \int \mathcal{L}(\mathbf{a}, \mathbf{z} \,|\, \boldsymbol{\theta}, m_Q) g(\boldsymbol{\theta} \,|\, m_Q) d\boldsymbol{\theta}$$

Assessing goodness-of-fit

- Goodness-of-fit diagnostics $\Rightarrow$ mostly computational, visualization based
- Ex: French political blog network from October 2006 [Kolaczyk'17]
  - $\Rightarrow$ We fit an SBM using variational MLE (**mixer** in R)



- Optimal value $\hat{Q} = 12$, but $Q \in [8, 12]$ reasonable (9 political parties)
  - $\Rightarrow$ Permuted adjacency shows group structure (room for merging few)
- Relatively good fit of the degree distribution

Extensions of SBMs

Degree-corrected SBMs

- Communities with broad degree distributions

B. Karrer B and M. E. Newman, "Stochastic blockmodels and community structure in networks," *Physical Review E.*, vol. 83, 2011

Mixed-membership SBMs

- Nodes may belong only partially to more than one class

E. M. Airoldi, "Mixed membership stochastic blockmodels," *J. Machine Learning Research*, vol. 9, 2008

Hierarchical SBMs

- Hierarchical clustering meets SBMs

A. Clauset et al, "Hierarchical structure and the prediction of missing links in networks," *Nature*, vol. 453, 2008

Roadmap

Random graph models

Small-world models

Network-growth models

Exponential random graph models

Latent network models

Random dot product graphs

Random dot product graphs

- Consider a latent space $\mathcal{X}_d \subset \mathbb{R}^d$ such that for all

$$\mathbf{x}, \mathbf{y} \in \mathcal{X}_d \quad \Rightarrow \quad \mathbf{x}^\top \mathbf{y} \in [0, 1]$$

$\Rightarrow$ Inner-product distribution $F : \mathcal{X}_d \mapsto [0, 1]$

- Random dot product graphs (RDPGs) are defined as follows:

$$\mathbf{x}_1, \ldots, \mathbf{x}_{N_v} \overset{\text{i.i.d.}}{\sim} F,$$
$$A_{ij} \,\big|\, \mathbf{x}_i, \mathbf{x}_j \sim \text{Bernoulli}(\mathbf{x}_i^\top \mathbf{x}_j)$$

for $1 \leq i, j \leq N_v$, where $A_{ij} = A_{ji}$ and $A_{ii} \equiv 0$

- A particularly tractable latent position random graph model

$\Rightarrow$ Vertex positions $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_v}]^\top \in \mathbb{R}^{N_v \times d}$

S. J. Young and E. R. Scheinerman, "Random dot product graph models for social networks," *WAW*, 2007

Connections to other models

- RDPGs ecompass several other classic models for network graphs

  Ex: Recover Erdös-Renyi $G_{N_v,p}$ graphs with $d = 1$ and $\mathcal{X}_d = \{\sqrt{p}\}$

  Ex: Recover SBM random graphs by constructing F with pmf

$$\mathrm{P}\left(\mathbf{X} = \mathbf{x}_q\right) = \alpha_q, \quad q = 1, \ldots, Q$$

  after selecting d and $\mathbf{x}_1, \ldots, \mathbf{x}_Q$ such that $\pi_{qr} = \mathbf{x}_q^\top \mathbf{x}_r$

- Approximation results for SBMs justify the expressiveness of RDPGs
- RDPGs are special cases of latent position models [Hoff et al'02]

$$A_{ij} \,\big|\, \mathbf{x}_i, \mathbf{x}_j \sim \text{Bernoulli}(\kappa(\mathbf{x}_i, \mathbf{x}_j))$$

  $\Rightarrow$ Approximate these accurately for large enough d [Tang et al'13]

Estimation of latent positions

- **Q:** Given G from an RDPG, find the 'best' $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_v}]^\top$?
- MLE is well motivated but it is intractable for large N_v

$$\hat{\mathbf{X}}_{ML} = \underset{\mathbf{X}}{\operatorname{argmax}} \prod_{i<j} (\mathbf{x}_i^\top \mathbf{x}_j)^{A_{ij}} (1 - \mathbf{x}_i^\top \mathbf{x}_j)^{1-A_{ij}}$$

- Instead, let $P_{ij} = \mathrm{P}\left((i,j) \in \mathcal{E}\right)$ and define $\mathbf{P} = [P_{ij}] \in [0,1]^{N_v \times N_v}$
  - $\Rightarrow$ RDPG model specifies that $\mathbf{P} = \mathbf{X}\mathbf{X}^\top$
  - $\Rightarrow$ **Key:** Observed $\mathbf{A}$ is a noisy realization of $\mathbf{P}$ $(\mathbb{E}[\mathbf{A}] = \mathbf{P})$
- Suggests a LS regression approach to find $\mathbf{X}$ s.t. $\mathbf{X}\mathbf{X}^\top \approx \mathbf{A}$

$$\hat{\mathbf{X}}_{LS} = \underset{\mathbf{X}}{\operatorname{argmin}} \|\mathbf{X}\mathbf{X}^\top - \mathbf{A}\|_F^2$$

A. Athreya et al, "Statistical inference on random dot product graphs: A survey," *J. Machine Learning Research*, 2018

Adjacency spectral embedding

- Since $\mathbf{A}$ is real and symmetric, can decompose it as $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$
  - $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_{N_v}]$ is the orthogonal matrix of eigenvectors
  - $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_{N_v})$, with eigvenvalues $\lambda_1 \geq \ldots \geq \lambda_{N_v}$
- Define $\hat{\boldsymbol{\Lambda}} = \text{diag}(\lambda_1^+, \ldots, \lambda_d^+)$ and $\hat{\mathbf{U}} = [\mathbf{u}_1, \ldots, \mathbf{u}_d]$ $(\lambda^+ := \max(0, \lambda))$
- Best rank-d, positive semi-definite (PSD) approx. of $\mathbf{A}$ is $\hat{\mathbf{P}} := \hat{\mathbf{U}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{U}}^\top$

  $\Rightarrow$ Adjacency spectral embedding (ASE) is $\hat{\mathbf{X}}_{ASE} = \hat{\mathbf{U}}\hat{\boldsymbol{\Lambda}}^{1/2}$ since

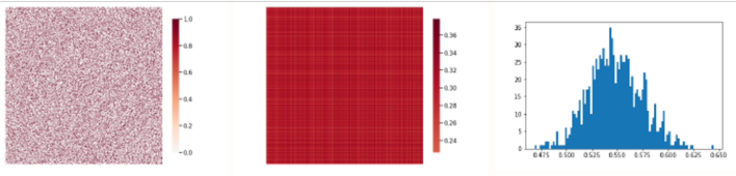
$$\mathbf{A} \approx \hat{\mathbf{U}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{U}}^\top = \hat{\mathbf{U}}\hat{\boldsymbol{\Lambda}}^{1/2}\hat{\boldsymbol{\Lambda}}^{1/2}\hat{\mathbf{U}}^\top = \hat{\mathbf{X}}_{ASE}\hat{\mathbf{X}}_{ASE}^\top$$

- **Q:** Is the solution unique? Nope, inner-products are rotation invariant

$$\mathbf{P} = \mathbf{X}\mathbf{W}(\mathbf{X}\mathbf{W})^\top = \mathbf{X}\mathbf{X}^\top, \quad \mathbf{W}\mathbf{W}^\top = \mathbf{I}_d$$

  $\Rightarrow$ RDPG embedding problem is identifiable modulo rotations

- Ex: Erdös-Renyi graph $G_{1000,0.3}$, realization of $\mathbf{A}$ (left)

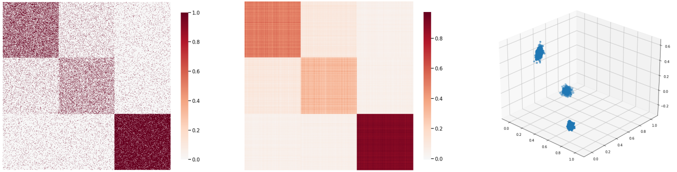


- For $d = 1$ we compute the ASE $\hat{\mathbf{x}}_{ASE}$ and plot $\hat{\mathbf{x}}_{ASE}\hat{\mathbf{x}}_{ASE}^\top$ (center)
  - $\Rightarrow$ Approximates well the constant matrix $\mathbf{P} = 0.3 \times \mathbf{1}\mathbf{1}^\top$
  - $\Rightarrow$ Supported by histogram of entries in $\hat{\mathbf{x}}_{ASE}$ (right, $\sqrt{p} = 0.547$)
- Consistency and limiting distribution results for ASEs available

Embedding an SBM graph

- Ex: SBM with $N_v = 1500$, $Q = 3$ and mixing parameters

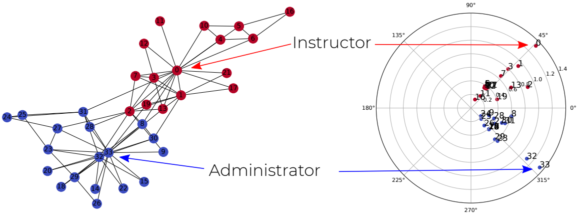
$$\boldsymbol{\alpha} = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \quad \boldsymbol{\Pi} = \begin{bmatrix} 0.5 & 0.1 & 0.05 \\ 0.1 & 0.3 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{bmatrix}$$



- Sample adjacency $\mathbf{A}$ (left), $\hat{\mathbf{X}}_{ASE}\hat{\mathbf{X}}_{ASE}^{\top}$ (center), rows of $\hat{\mathbf{X}}_{ASE}$ (right)
- Use embeddings to bring to bear geometric methods of analysis

Interpretability of the embeddings

- Ex: Zachary's karate club graph with $N_v = 34$, $N_e = 78$ (left)



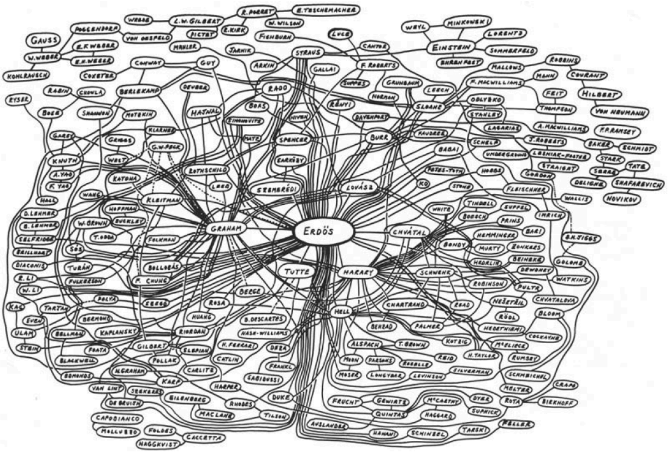
- Node embeddings (rows of $\hat{\mathbf{X}}_{ASE}$) for $d = 2$ (right)
  - Club's administrator ($i = 0$) and instructor ($j = 33$) are orthogonal
- Interpretability of embeddings a valuable asset for RDPGs

  $\Rightarrow$ Vector magnitudes indicate how well connected nodes are

  $\Rightarrow$ Vector angles indicate positions in latent space

Mathematicians collaboration graph

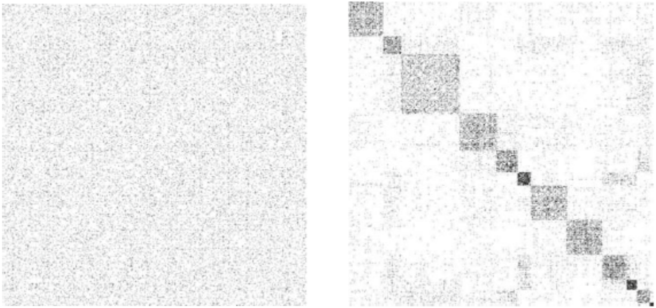
- Ex: Mathematics collaboration network centered at Paul Erdös



- Most mathematicians have an Erdös number of at most 4 or 5
  - ⇒ Drawing created by R. Graham in 1979

Mathematicians collaboration graph

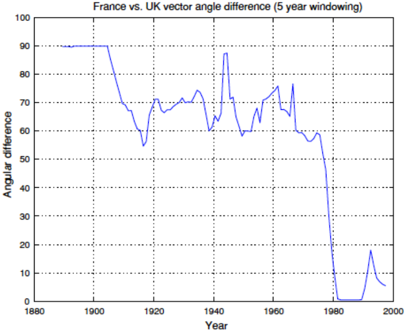
- Coauthorship graph G, $N_v = 4301$ nodes with Erdös number ≤ 2
  - $\Rightarrow$ No discernible structure from the adjacency matrix $\mathbf{A}$ (left)



- Community structure revealed after row-colum permutation (right)
  - (i) Obtained the ASE $\hat{\mathbf{X}}_{ASE}$ for the mathematicians
  - (ii) Performed angular k-means on $\hat{\mathbf{X}}_{ASE}$'s rows [Scheinerman-Tucker'10]

International relations

- Ex: Dynamic network G_t of international relations among nations
  - $\Rightarrow$ Nations $(i, j) \in \mathcal{E}_t$ if they have an alliance treaty during year t



- Track the angle between UK and France's ASE from 1890-1995
  - Orthogonal during the late 19th century
  - Came closer during the wars, retreat during Nazi invasion in WWII
  - Strong alignment starts in the 1970s in the run up to the EU

Closing remarks and extensions

- Neglected the constraint $[\hat{\mathbf{X}}_{ASE}\hat{\mathbf{X}}_{ASE}^\top]_{ii} = 0$. Fix via iterative algorithm

  E. R. Scheinerman and K. Tucker, "Modeling graphs using dot product representations," *Comput. Stat.*, vol. 25, pp. 1-16, 2010

- Assumed **A** to be PSD. Extension known as generalized RDPG

  P. Rubin-Delanchy et al, "A statistical interpretation of spectral embedding: The generalised random dot product graph," *arXiv:1709.05506*, 2017

- RDPG variants to model weighted and directed graphs possible

  F. Larroca et al, "Change point detection in weighted and directed random dot product graphs," *EUSIPCO*, 2021

- Host of applications in testing, clustering, change-point detection, . . .

Glossary

- Network graph model
- Random graph models
- Configuration model
- Matching algorithm
- Switching algorithm
- Model-based estimation
- Assessing significance
- Reference distribution
- Network motif
- Small-world network
- Decentralized search
- Watts-Strogatz model
- Time-evolving network
- Network-growth models
- Preferential attachment
- Barabási-Albert model
- Copying models
- Exponential family
- Exponential random graph models
- Configurations
- Network statistic
- Homogeneity
- Markov random graphs