ST5225: Statistical Analysis of Networks
Lecture 12: Review

WANG Wanjie
staww@nus.edu.sg

Department of Statistics and Applied Probability
National University of Singapore (NUS)

Saturday 21 April, 2018

Final Exam

- Time: 2018.5.8, 13:00-15:00
- Location: S16-04-30/41
- Requirement: Closed Book; One double-sided help sheet; Non-programming calculator
- Questions: Similar as assignments; No coding required
- Office hour: 2-4m on Thursdays
- Coverage: Lectures 1-10 (Lecture 11: latent position model is not covered)

Basics about networks

- Network: nodes, links, additional info.
- Directed/Undirected
  - Relationship between directed graph and undirected graph

- Simple Graph: no multiple edges, self loops
- Neighbors: Given node A, its neighbors are the nodes adjacent to it.
  - Neighbors of a set S: $N(S)$
- Paths: a way to get from one node to another along edges
- Distance
  - Geodesic distance on the graph: minimal distance of paths between nodes i and j
  - Euclidean distance defined according to their neighbors
  - etc.

Structures of a graph

- Sub-graph: induced sub-graph
- Connected/Unconnected
  - There is a path for every pair of nodes
  - components, giant component
  - Directed graph: strongly connected, weakly connected
  - Bowtie structure
- Complete
  - Clique
  - Maximal clique
  - k-core

Mathematics about a graph

- Adjacency matrix

$$A_{ij} = \begin{cases} 1, & \text{if } (i,j) \in E \\ 0, & \text{otherwise} \end{cases}$$

  Adjacency matrix uniquely decides a network.

- Degree
  - Undirected graph: number of neighbors / incident edges / row(column) sums of the adjacency matrix
  - Directed graph: out-degree, in-degree, total degree
- Degree distribution: $f_d \propto d^{-\alpha}$
- Diameter
  - One kind of centrality
  - Rank pages: Authority-hub algorithm

- Induced-subgraph sampling
- Incident-subgraph sampling
- Snowball sampling: $S \to S \cup N(S) \to$the subgraph induced by $S \cup N(S)$
- Respondent-driven sampling: the number of neighbors is restricted
- Trace-route sampling: a set of starting points and a set of ending points.
- Choice of sampling methods depend on the research
- Horvitz-Thompson estimator

$$\hat{\mu}_{HT} = \frac{1}{n} \sum_{i \in S} \frac{X_i}{\pi_i},$$

  where π_i is the inclusion probability of X_i
- Unbiased estimator. Yet it cannot solve all the problems

Centrality

- Closeness: $\frac{1}{\frac{1}{n-1}\sum_{j\neq i} d(i,j)} = \frac{n-1}{\sum_{j\neq i} d(i,j)}$.
- Note: Normalization
- Betweenness
  - $\sigma(u,v)$: number of shortest paths between u and v
  - $\sum_{u\neq i, v\neq i, u\neq i} \frac{\sigma(u,v)}{\sigma(u,v|i)}$
  - For undirected graph, we consider all the pairs (u,v) *without* the ordering. For directed graph, we consider all the pairs.
  - Edge betweenness – one graph partition method
- Eigenvector centrality
  - The eigenvector corresponding to the largest eigenvalue of the adjacency matrix A
  - For directed graphs, consider AA^T or A^TA.
- Comparison of three measurements of centrality

Cohesion

- Cliques; maximal cliques
- (maximal) k-core; coreness
- k-vertex/edge-connectivity
- Local density
  - undirected graph: $\frac{|E|}{|V|(|V|-1)/2}$;
  - directed graph: $\frac{|E|}{|V|(|V|-1)}$
  - density of a node
- transitivity / clustering coefficient
  - Node: $\frac{\#\text{triangles } v \text{ falls into}}{\#\text{connected triples that both edges are incident to } v}$
  - Graph: $\frac{3\times\#\text{triangles in the graph}}{\#\text{connected triples in the graph}}$
  - prob that u connects with v, given both u and v are connected with w

Graph Partition

- Partition of nodes
- Edge betweenness
- hierarchical clustering
  - Evaluation of a partition
    - (dis)similarity between nodes → linkage to evaluate a partition
    - Linkages: complete/single/average/etc.
    - Modularity

$$\sum_{k=1}^{K} [f_{kk} - f_{k+} f_{+k}]$$

  - For each step, find the proper decomposition/combination of the sets, so that the valuation of the partition is high.
  - Dendrogram
  - Cut the dendrogram for proper result

Matching Markets

- Bipartite network: a special type of network
  - Balanced bipartite network
- Perfect matching
  - Constricted set: $|S| > |N(S)|$
- Optimal assignment
  - Each buyer has valuations on every product (seller)
  - Maximize the total evaluation
- Market-clearing prices
  - Prices and payoffs
  - With the payoff, there is an optimal assignment
  - Preferred-seller graph
  - With the preferred-seller graph, there might be a perfect matching
  - If there is a perfect matching, that perfect matching must be the optimal assignment.
  - We call that set of prices as market-clearing prices

Market-clearing prices

- Market-clearing prices are not unique. It always exist.
- With the initial prices, we can find a series of market-clearing prices

  Start: each item has price 0; each buyer assigns a value to each item

  1. Assume smallest value is 0; if not, scale the price so that the smallest is 0
  2. Construct the preferred-seller graph and check if there is a perfect matching
  3. If yes, done
  4. If not, find a constricted set of buyers S
  5. Each seller in $N(S)$ increases the price by 1
  6. check if the smallest price is 0, if not, subtract the same amount of each price so that the smallest is 0
  7. go back to Step 1.

- Power of nodes on a network
- Nash bargaining solution
  - A pair of nodes i and j, where x has outside option as x, and j has outside option as y.
  - The rational outcome is

$$A : x + \frac{1 - x - y}{2}, \qquad B : y + \frac{1 - x - y}{2},$$

    and there is no transaction if $x + y > 1$.
  - Note that x is the portion A gets if A trades with outside nodes.
- Stable outcome: check the instability for every edge that does not belong to the matching.
- Balanced stable outcome: check every edge, make sure that they satisfy the Nash bargaining solution
- Stable outcome may not exist

Word Wide Web network

- Definition of WWW network: webpages and hyperlinks
- Authority-Hub algorithm
  1. Each page v has two scores, $auth(v)$, $hub(v)$
  2. Start with $hub(v) = 1$ for each v
  3. Repeat
     - Normalize $hub(v)$ so that $\sum_v hub(v) = 1$
     - For each v, update $auth(v) = \sum_{u,(u,v)\in E} hub(u)$
     - Normalize $auth(v)$ so that $\sum_v auth(v) = 1$
     - For each v, update $hub(v) = \sum_{u,(v,u)\in E} auth(u)$
  4. Output the result according to the authorities
- Mathematical analysis:

$$H^{(k+1)} = A \times Auth^{(k+1)} = A \times A^T \times H^{(k)} = AA^T H^{(k)},$$
$$Auth^{(k+1)} = A^T \times H^{(k)} = A^T \times A \times Auth^{(k)} = A^T A \times Auth^{(k)}$$

- Consider k steps. Output the result after k repetitions, say, $k = 2$
- Or, repeat the procedure many times, until the scores converge. Then Hub score (authority score, respectively) converges to the top eigenvector of AA^T ($A^T A$, respectively).

Example

Page Rank

- Page Rank: only one score for each node
  1. Each page starts with PageRank of $1/|V|$.
  2. Each page pass the scores to the other pages. For each node v, if the PageRank is $r(v)$ and the out-degree is K, then it passes $r(v)/K$ to each neighbour.
  3. Update $r(v)$ for each node to be the sum of the scores received
  4. Repeat the "passing-receiving-updating" procedure for k steps.
- Random walk on the network: $r^{(k)} = rP^k$, where r is the initial PageRank score, $r^{(k)}$ is the PageRank score after k steps, and P is the transition probability matrix, where

$$P_{ij} = \begin{cases} \frac{1}{d_i^{out}} = \frac{1}{\sum_k A_{ik}}, & \text{if } (i,j) \in E \\ 0, & \text{otherwise} \end{cases}$$

- Scaled PageRank:
  1. After each repetition, scale the PageRank to be $r(v) \times s$
  2. Add $\frac{1-s}{|V|}$ to each node

Example

Advertisement

- Advertisement
  - How to set the price for advertisements in search engines
  - clickthrough rate, revenue per click, valuation, matching market
  - Fake spots/advertisers
- Review of statistical notions
  - Model: a set of dist.
  - probability density function
  - joint prob. density function, likelihood function
  - Maximum likelihood estimate
  - Common dist.: Bernoulli dist, Binomial dist., uniform dist., normal dist., Poisson dist.

Random Graph

- $|V|$ is given, $(i,j) \overset{i.i.d.}{\sim} Bernoulli(p)$
  - Undirected graph: $A_{ij} = A_{ji} \sim Bernoulli(p)$
  - Directed graph: $A_{ij} \sim Bernoulli(p)$, $A_{ji} \sim Bernoulli(p)$, independent
- Likelihood: $p^{|E|}(1-p)^{\binom{|V|}{2}-|E|}$ (undirected), $p^{|E|}(1-p)^{|V|(|V|-1)-|E|}$ (directed),
- MLE: $|E|/\binom{|V|}{2}$ (undirected), $|E|/|V|(|V|-1)$ (directed)
- Degree dist: $Binomial(n,p)$
- Parameterization
  - Fix p
  - Fix $\lambda = np$
- drawbacks of the model: few triangles, no clustering structure, degree dist.

Stochastic Block Model

- Community labels are given
  - Each node has a label ℓ_i, indicating which community it belongs to. The prob. of an edge depends on ℓ_i and ℓ_j
  - Probability matrix: $\Pi B \Pi^T - diag(\Pi B \Pi^T)$
  - Likelihood:
    $L = \prod_{r \neq s} b_{rs}^{e_{rs}} (1 - b_{rs})^{n_r n_s - e_{rs}} \times \prod_r b_{rr}^{e_{rr}} (1 - b_{rr})^{\binom{n_r}{2} - e_{rr}}$
  - MLE: $\hat{b}_{rs} = \frac{e_{rs}}{n_r n_s}, \qquad \hat{b}_{rr} = \frac{e_{rr}}{\binom{n_r}{2}}$
  - Degree dist: $\sum_{k=1}^{K} Binomial(n_k - \delta_{\ell_i,k}, b_{\ell_i,k})$.
- Community labels are unknown
  - If the labels are unknown, we model the labels as multinomial dist., and have new likelihood function
  - MLE does not have explicit solution in this case

Exponential Random Graph Model

- Sufficient Statistics
  - Statistics: a function of data, $T(X)$
  - Sufficient statistics: $P(X|T(X), \theta) = P(X|T(X))$, given the sufficient statistics, the conditional probability does not depend on the unknown parameters θ
  - Factorization Theorem: $P)\theta(x) = h(x)g(\theta, T(x)) \Leftrightarrow T(x)$ is sufficient stat. for θ
- Exponential family distributions
  - Definition: $f_\theta = h(x)g(\theta)\exp\{\sum_{i=1}^{d} \theta_i T_i(x)\}$
  - Relationship between suff. stat. and exponential family dist.
- Exponential Random Graph Model
  - Definition:

$$f_\theta(A) = h(\theta)\exp\{\sum_{i=1}^{d} \theta_i T_i(A)\}, \qquad h(\theta) = 1/\sum_{x} \exp\{\sum_{i=1}^{d} \theta_i T_i(x)\},$$

    where $T_i(A)$ are sufficient statistics, and $h(\theta)$ is the normalizing constant
  - Examples: Random graph model, stochastic block model, degree correction model, etc.

Exponential Random Graph Model

- Construction of ERGM
  - Pick d functions of the graph, $T_1(A), T_2(A), \cdots, T_d(A)$
  - The density function is $f_\theta(A) = h(\theta)\exp\{\sum_{i=1}^d \theta_i T_i(A)\}$, where $h(\theta) = 1/\sum_x \exp\{\sum_{i=1}^d \theta_i T_i(x)\}$.
  - Known parameters, apply gibbs sampling to draw graphs
  - Unknown parameters and observed data, apply stochastic approximation
- Log odds of edge prob.

$$\log \frac{P(A_{ij}=1)}{1-P(A_{ij}=1)} = \sum_{k=1}^{d} (T_k(A_{+ij}) - T_k(A_{-ij}))\theta_k.$$

- MLE: MLE should satisfy that $T_i(A) = E_{\hat{\theta}}[T_i]$, any $1 \le i \le d$.
- Example: $p1$ model