Kronecker Graphs
Models of Networks

- **What is the goal of modeling networks?**
  - Discover structural properties of networks
    - Small-world, Edge clustering, Heavy-tailed degrees
  - Find a model that gives graphs with such properties
    - Erdos-Renyi, Watts-Strogatz, Barabasi-Albert model

- **Today’s lecture:**
  - Can we have a model that reproduces all properties?
  - Can we fit the model to a network and accurately reproduce it?

- **Kronecker Graphs** [Leskovec et al. ‘05]
How can we think of network structure recursively? **Intuition: Self-similarity**

- **Object is similar to a part of itself:** the whole has the same shape as one or more of the parts

- **Mimic recursive graph/community growth**

- **Kronecker graph** is a way of generating self-similar matrices
Kronecker: Graph Growth

Intermediate stage

\[
\begin{bmatrix}
1 & 1 & 0 \\
1 & 1 & 1 \\
0 & 1 & 1 \\
\end{bmatrix}
\quad \text{Initiator graph}
\]

\[
\begin{bmatrix}
K_1 & K_1 & 0 \\
K_1 & K_1 & K_1 \\
0 & K_1 & K_1 \\
\end{bmatrix}
\quad \text{After the growth phase}
\]

\[
K_2 = K_1 \otimes K_1
\]
How can we think of network structure recursively?

\[
K_1 = \begin{pmatrix}
a & b \\
c & d \\
\end{pmatrix}
\]
Kronecker Graph

- Kronecker graphs:
  - A recursive model of network structure

\[ K_2 = K_1 \otimes K_1 \]

3 x 3  \quad 9 x 9  \quad 27 x 27 \text{ adjacency matrix}
Kronecker Product: Definition

- **Kronecker product** of matrices $A$ and $B$ is given by

$$C = A \otimes B = \begin{pmatrix}
    a_{1,1}B & a_{1,2}B & \ldots & a_{1,m}B \\
    a_{2,1}B & a_{2,2}B & \ldots & a_{2,m}B \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{n,1}B & a_{n,2}B & \ldots & a_{n,m}B
\end{pmatrix}
$$

$N \times M \times K \times L \rightarrow N^*K \times M^*L$

- Define a Kronecker product of two **graphs** as a Kronecker product of their **adjacency matrices**
Kronecker Graphs

- **Kronecker graph**: a growing sequence of graphs by iterating the Kronecker product:

\[
K_1^{[k]} = K_k = \underbrace{K_1 \otimes K_1 \otimes \ldots K_1}_{k \text{ times}} = K_{k-1} \otimes K_1
\]

- **Note**: One can easily use multiple initiator matrices \((K_1', K_1'', K_1''')\) (even of different sizes)
Kronecker Initiator Matrices

Initiator $K_1$

$K_1$ adjacency matrix

$K_3$ adjacency matrix
Kronecker Graphs: First Fact

\[ K_{1}^{[k]} = K_{k} = K_{1} \otimes K_{1} \otimes \ldots \otimes K_{1} = K_{k-1} \otimes K_{1} \]

\[ k \text{ times} \]

First fact about Kronecker Graphs!

- For \( K_{1} \) on \( N_{1} \) nodes and \( E_{1} \) edges
  \( K_{k} \) (\( k^{th} \) Kronecker power of \( K_{1} \)) has:
  - \( N_{1}^{k} \) nodes
  - \( E_{1}^{k} \) edges

- So, we get the densification power-law!
  - \( E(t) \propto N(t)^{\alpha} \), What is \( \alpha \)?
  - \( \alpha = \frac{\log(E(t))}{\log(N(t))} = \frac{\log(E_{1})}{\log(N_{1})} \)
Properties of deterministic Kronecker graphs (can be proved!)

- Properties of static networks:
  - Power-Law like Degree Distribution
  - Power-Law eigenvalue and eigenvector distribution
  - Constant Diameter (next)

- Properties of evolving networks:
  - Densification Power Law (just proved)
  - Shrinking/Stabilizing Diameter (for Stochastic Kronecker graphs)
**Observation:** Edges in Kronecker graphs:

\[ \text{Edge } (X_{ij}, X_{kl}) \in G \otimes H \]

iff \((X_i, X_k) \in G \text{ and } (X_j, X_l) \in H\)

where \(X\) are appropriate nodes in \(G\) and \(H\)

**Why?**

- An entry in matrix \(G \otimes H\) is a multiplication of entries in \(G\) and \(H\).
**Theorem:** **Constant diameter:** If graphs $G, H$ have diameter $d$ then $G \bowtie H$ has diameter $d$

**What is distance between nodes $u, v$ in $G \otimes H$?**

- Let $u = [a, b], v = [a', b']$ (using the notation from the last slide)
  - then edge $(u, v)$ in $G \otimes H$ if $(a, a') \in G$ and $(b, b') \in H$
- So, path $a$ to $a'$ in $G$ is less $d$ steps: $a_1, a_2, a_3, ..., a_d$
- And path $b$ to $b'$ in $H$ is less $d$ steps: $b_1, b_2, b_3, ..., b_d$
- Then: Edge $([a_1, b_1], [a_2, b_2])$ is in $G \otimes H$
- So it takes $< d$ steps to get from $u$ to $v$ in $G \bowtie H$

**Consequence:**

- If $K_1$ has diameter $d$ then graph $K_k$ also has diameter $d$
Stochastic Kronecker Graphs

- Create $N_1 \times N_1$ probability matrix $\Theta_1$
- Compute the $k^{th}$ Kronecker power $\Theta_k$
- For each entry $p_{uv}$ of $\Theta_k$ include an edge $(u, v)$ in $K_k$ with probability $p_{uv}$

**Example:**

\[
\Theta_1 = \begin{pmatrix}
0.5 & 0.2 \\
0.1 & 0.3
\end{pmatrix}
\]

\[
\Theta_2 = \Theta_1 \otimes \Theta_1 = \begin{pmatrix}
0.25 & 0.10 & 0.10 & 0.04 \\
0.05 & 0.15 & 0.02 & 0.06 \\
0.05 & 0.02 & 0.15 & 0.06 \\
0.01 & 0.03 & 0.03 & 0.09
\end{pmatrix}
\]

**Probability of edge $p_{uv}$**
Stochastic Kronecker Graphs

What is known about Stochastic Kronecker?

- **Undirected** Kronecker graph model with:
  - **Connected**, if:
    - $b + c > 1$
  - **Connected component of size $\Theta(n)$**, if:
    - $(a + b)(b + c) > 1$
  - **Constant diameter**, if:
    - $b + c > 1$
  - **Not searchable** by a decentralized algorithm

\[ \Theta_1 = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \]

$\Omega_1 = \begin{pmatrix} a > b > c \end{pmatrix}$
How do we generate an instance of a stochastic Kronecker graph?

<table>
<thead>
<tr>
<th>Probability of edge $p_{uv}$</th>
<th>0.25</th>
<th>0.10</th>
<th>0.10</th>
<th>0.04</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.15</td>
<td>0.02</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>0.02</td>
<td>0.15</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
<td>0.09</td>
<td></td>
</tr>
</tbody>
</table>

Flip biased coins

<table>
<thead>
<tr>
<th>1</th>
<th>1</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Need to flip $n^2$ coins!!
Way too slow!!

Is there a faster way? YES!

Idea: Exploit the recursive structure of Kronecker graphs

“Drop” edges one by one
**Generation of Kronecker Graphs**

- **A faster way to generate Kronecker graphs**

  \[ \Theta = \begin{array}{cc}
  a & b \\
  c & d 
  \end{array} \]

  \[ \mathcal{G} \text{ on } n = k^2 \text{ nodes} \]

  \[
  \begin{align*}
  L_{ij} &= \Theta_{ij} / (\sum_{op} \Theta_{op}) \\
  \text{For } m = 1 \ldots k \\
  \text{start with } x = 0, y = 0 \\
  \text{Pick an row/column } (i, j) \text{ with prob. } L_{ij} \\
  \text{Descent into quadrant } (i, j) \text{ at level } m \text{ of } \mathcal{G} \\
  \text{This means: } x &+\!\!\!\!\!\!\!\!\!= i \cdot 2^{k-m}, \ y &+\!\!\!\!\!\!\!\!\!= j \cdot 2^{k-m} \\
  \text{Add an edge } G[x, y] &= 1
  \end{align*}
  \]

- **Generate an edge graph \( \mathcal{G} \) on \( n = k^2 \) nodes**
  
  \[
  \begin{array}{cccc}
  \mathcal{V}_1 & \mathcal{V}_2 & \mathcal{V}_3 & \mathcal{V}_4 \\
  a \cdot a & a \cdot b & b \cdot a & b \cdot b \\
  a \cdot c & a \cdot d & b \cdot c & b \cdot d \\
  c \cdot a & c \cdot b & d \cdot a & d \cdot b \\
  c \cdot c & c \cdot d & d \cdot c & d \cdot d \\
  \end{array}
  \]

  \[
  K_2 = K_1 \otimes K_1
  \]

11/7/2012

How to estimate $\Theta$ given a $G$?

- **KronFit**: Maximum likelihood estimation
- Given real graph $G$
- Find Stochastic Kronecker initiator $\Theta$ which

$$\Theta = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$\arg \max_{\Theta} P(G | \Theta)$$

- **To solve this we need to:**
  - Efficiently calculate $P(G | \Theta)$
  - Then maximize over $\Theta$ (e.g., using gradient descent)
Given \( G \) and \( \Theta \) we calculate likelihood that \( \Theta \) generated \( G \): \( P(G|\Theta) \)

\[
P(G|\Theta) = \prod_{(u,v) \in G} \Theta_k[u,v] \prod_{(u,v) \not\in G} (1 - \Theta_k[u,v])
\]
Nodes are unlabeled

Graphs \( G' \) and \( G'' \) should have the same probability \( P(G'|\Theta) = P(G''|\Theta) \)

One needs to consider all node correspondences \( \sigma \)

\[
P(G|\Theta) = \sum_{\sigma} P(G|\Theta, \sigma)P(\sigma)
\]

All correspondences are a priori equally likely

There are \( O(n!) \) correspondences

\[
P(G'|\Theta) = P(G''|\Theta)
\]
Challenge 2: Calculating $P(G | \Theta, \sigma)$

- Assume that we solved the node correspondence problem
- Calculating

$$P(G | \Theta) = \prod_{(u,v) \in G} \Theta_k[u,v] \prod_{(u,v) \notin G} (1 - \Theta_k[u,v])$$

- Takes $O(n^2)$ time!
**Node correspondence:**

- Permutation $\sigma$ defines the mapping
- Randomly search over $\sigma$ to find good mappings

**Calculating the likelihood $P(G|\Theta,\sigma)$**

- Calculate likelihood of empty graph (G with 0 edges)
- Correct it for edges that we observe in the graph
Solution 1: Node correspondence

- Log-likelihood

\[
    l(\Theta) = \log \sum_{\sigma} P(G | \Theta, \sigma) P(\sigma)
\]

- Gradient of log-likelihood

\[
    \frac{\partial}{\partial \Theta} l(\Theta) = \sum_{\sigma} \frac{\partial \log P(G | \sigma, \Theta)}{\partial \Theta} P(\sigma | G, \Theta)
\]

- Sample the permutations from \( P(\sigma | G, \Theta) \) and average the gradients

See Leskovec-Faloutsos, ICML '07 for details
Solution 1: Node correspondence

- **Metropolis sampling:**
  - Start with a random permutation $\sigma$
  - $\sigma' = $ swap two elements in permutation $\sigma$
  - Accept the new permutation $\sigma'$
    - If new permutation is better (gives higher likelihood)
    - Else accept with prob. proportional to the ratio of likelihoods
      (no need to calculate the normalizing constant!)

\[
\frac{P(\sigma' | G, \Theta)}{P(\sigma | G, \Theta)}
\]

See Leskovec-Faloutsos, ICML '07 for details
Sampling node labelings (2)

\[\sigma^{(0)} := (1, \ldots, N)\]

repeat
  Draw \(j\) and \(k\) uniformly from \((1, \ldots, N)\)
  \[\sigma^{(i)} := \text{SwapElements}(\sigma^{(i-1)}, j, k)\]
  Draw \(u\) from \(U(0, 1)\)
  if \(u > \frac{P(\sigma^{(i)}|G, \Theta)}{P(\sigma^{(i-1)}|G, \Theta)}\) then
    \[\sigma^{(i)} := \sigma^{(i-1)}\]
  end if
  \(i = i + 1\)
until \(\sigma^{(i)} \sim P(\sigma|G, \Theta)\)
return \(\sigma^{(i)}\)

- Need to efficiently calculate the likelihood ratios
- But the permutations \(\sigma^{(i)}\) and \(\sigma^{(i+1)}\) only differ at 2 positions
- So we only traverse to update 2 rows (columns) of \(\Theta_k\)
- We can evaluate the likelihood ratio efficiently
Solution 2: Calculating $P(G|\Theta,\sigma)$

- Problem:
  - Calculating naively $P(G|\Theta,\sigma)$ takes $O(N^2)$

- Idea:
  - First calculate likelihood of empty graph, a graph with 0 edges
  - Correct the likelihood for edges that we observe in the graph

- By exploiting the structure of Kronecker product we obtain closed form for likelihood of an empty graph

See Leskovec-Faloutsos, ICML ‘07 for details
We approximate the likelihood:

\[ l(\Theta) \approx l_e(\Theta) + \sum_{(u,v) \in G} -\log(1 - \Theta_k[\sigma_u, \sigma_v]) + \log(\Theta_k[\sigma_u, \sigma_v]) \]

- The sum goes only over the edges
- Evaluating \( P(G|\Theta, \sigma) \) takes \( O(e) \) time
- Real graphs are sparse, \( e \ll n^2 \)

See Leskovec-Faloutsos, ICML '07 for details
Real graphs are sparse so we first calculate likelihood of empty graph

Probability of edge \((i,j)\) is in general \(p_{ij} = \theta_1^a \theta_2^b \theta_3^c \theta_4^d\)

By using Taylor approximation to \(p_{ij}\) and summing the multinomial series we obtain:

\[
l_e(\Theta) = \sum_{i,j=1}^{N} \log(1 - p_{ij}) \approx -\left( \sum_{i,j=1}^{N_1} \theta_{i,j} \right)^k - \frac{1}{2} \left( \sum_{i,j=1}^{N_1} \theta_{i,j}^2 \right)^k
\]

We approximate the likelihood:

\[
l(\Theta) \approx l_e(\Theta) + \sum_{(u,v) \in G} -\log(1 - \Theta_k[\sigma_u, \sigma_v]) + \log(\Theta_k[\sigma_u, \sigma_v])
\]

Taylor approximation \(\log(1-x) \sim -x - 0.5 x^2\)
Experimental setup

- Given real graph $G$
- Gradient descent from random initial point
- Obtain estimated parameters $\Theta$
- Generate synthetic graph $K$ using $\Theta$
- Compare properties of graphs $G$ and $K$

Note:

- We do not fit the graph properties themselves
- We fit the likelihood and then compare the properties
Real and Kronecker are very close:

$$\Theta_1 = \begin{bmatrix} 0.99 & 0.54 \\ 0.49 & 0.13 \end{bmatrix}$$
What do estimated parameters tell us about the network structure?
What do estimated parameters tell us about the network structure?

\[ \Theta = \begin{pmatrix} 0.9 & 0.5 \\ 0.5 & 0.1 \end{pmatrix} \]
Small and large networks are very different:
Large scale network structure:

- Large networks are different from small networks and manifolds
- Nested Core-periphery
  - Recursive onion-like structure of the network where each layer decomposes into a core and periphery
Remember the SKG theorems:

- **Connected**, if \( b+c>1 \):
  - \( 0.55+0.15 > 1 \). No!

- **Giant component**, if \( (a+b) \cdot (b+c)>1 \):
  - \( (0.99+0.55) \cdot (0.55+0.15) > 1 \). Yes!

Real graphs are in the parameter region analogous to the giant component of an extremely sparse \( G_{np} \).
A Different Model: MAG Model
When modeling networks, what would we like to know?

- How to model the links in the network
- How to model the interaction of node attributes/properties and the network structure

Goal:

- A family of models of networks with node attributes
- The models are:
  1) Analytically tractable (prove network properties)
  2) Statistically interesting (can be fit to real data)
The MAG model

Node Features

- Democrat
- Male
- Out-going
- Democrat
- Female
- Shy

Network Structure

Homophily

- Democrat to Democrat: 0.9
- Male to Female: 0.2
- Out-going to Shy: 0.5

Heterophily

- Democrat to Male: 0.1
- Democrat to Female: 0.8
- Male to Democrat: 0.9
- Female to Male: 0.1

Core-periphery

- Democrat to Out-going: 0.9
- Male to Out-going: 0.9
- Out-going to Democrat: 0.5
- Out-going to Male: 0.2

Link Affinity

- Homophily: 0.9
- Heterophily: 0.2
- Core-periphery: 0.5

Network

\[ P(\text{ Democrat } \rightarrow \text{ Female }) = 0.9 \times 0.9 \times 0.5 \]
Each node has a set of **categorical attributes**

- Example:
  - Gender: Male, Female
  - Home country: US, Canada, Russia, etc.

How do node attributes influence link formation?

<table>
<thead>
<tr>
<th></th>
<th>( u )'s gender</th>
<th>( v )'s gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FEMALE</td>
<td>MALE</td>
</tr>
<tr>
<td>( u ) is friends with ( v )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEMALE</td>
<td>0.3</td>
<td>0.6</td>
</tr>
<tr>
<td>MALE</td>
<td>0.6</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Let the values of the *i*-th attribute for node \( u \) and \( v \) be \( a_i(u) \) and \( a_i(v) \)
- \( a_i(u) \) and \( a_i(v) \) can take values \( \{0, \ldots, d_i - 1\} \)

**Question:** How can we capture the influence of the attributes on link formation?

**Attribute link-affinity matrix** \( \Theta \)

\[
P(u, v) = \Theta[a_i(u), a_i(v)]
\]

Each entry of the attribute matrix captures the *probability of a link* between two nodes associated with the attributes of them.
Combining attributes

- How do we combine the effects of multiple attributes?
  - Multiply the probabilities from all attributes

\[
\alpha(u) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \alpha(v) = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \Theta_i = \begin{bmatrix} \alpha_1 & \beta_1 & \alpha_2 & \beta_2 & \alpha_3 & \beta_3 & \alpha_4 & \beta_4 \\ \beta_1 & \gamma_1 & \beta_2 & \gamma_2 & \beta_3 & \gamma_3 & \beta_4 & \gamma_4 \end{bmatrix}
\]

\[
P(u, v) = \alpha_1 \times \beta_2 \times \gamma_3 \times \alpha_4
\]
Link-Afinity Matrices offer **flexibility in modeling the network structure**:

- **Homophily**: love of the *same*
  - e.g., political parties, hobbies

- **Heterophily**: love of the *opposite*
  - e.g., genders

- **Core-periphery**: love of the *core*
  - e.g. extrovert personalities
**Multiplicative Attribute Graph**

- **Multiplicative Attribute Graph** $M(n, l, \vec{a}, \vec{\Theta})$:
  - A network contains $n$ nodes
  - Each node has $l$ categorical attributes
  - $a_i(u)$ represents the $i$-th attribute of node $u$
  - Each attribute $a_i(\cdot)$ is linked to a $d_i \times d_i$ attribute link-affinity matrix $\Theta_i$
  - Edge probability between nodes $u$ and $v$

$$P(u, v) = \prod_{i=1}^{l} \Theta_i[a_i(u), a_i(v)]$$
Each node $u$ has associated binary vector $A_u$

Think of as answered to a set of yes/no questions

**Probability of a link** between nodes $u$, $v$:

$$P(u, v) = \prod_{i=1}^{k} K(A_u(i), A_v(i))$$

$$K = \begin{bmatrix} 0 & 1 & 0 \\ a & b & 0 \\ c & d & 1 \end{bmatrix}$$

$v_2 = (0,1)$

$v_3 = (1,0)$

$P(v_2, v_3) = b \cdot c$

$K_2 = K_1 \otimes K_1$
Each node in a Kronecker graph has a node id (e.g. $0, \cdots, 2^l - 1$)

A binary representation of node id is its attribute vector in a MAG model

Then, the (stochastic) adjacency matrices of two models are equivalent

Example:

$$
K = \begin{pmatrix}
  a & b \\
  c & d
\end{pmatrix}
$$

$$
\begin{array}{c|c|c|c}
  & v_0 & v_1 & v_2 \\
\hline
v_0 & a & b & a \\
v_1 & c & d & b \\
v_2 & a & b & a \\
v_3 & c & d & c \\
\end{array}
$$

$$
\begin{align*}
a(v_1) &= [0 \ 1] \\
a(v_2) &= [1 \ 0] \\
P(v_1,v_2) &= b \cdot c
\end{align*}
$$
MAG can model global network structure!

MAG generates networks with similar properties as found in real-world networks:

- Unique giant connected component
- Densification Power Law
- Small diameter
- Heavy-tailed degree distribution
  - Either log-normal or power-law
Theorem 1: A unique giant connected component of size $\theta(n)$ exists in $M(n, l, \mu, \Theta)$ w.h.p. as $n \to \infty$ if $P(a_i(u) = 1) = \mu$

$$\left[\left(\mu\alpha + (1 - \mu)\beta\right)^\mu (\mu\beta + (1 - \mu)\gamma)^{1 - \mu}\right]^\rho \geq \frac{1}{2}$$

- Simulation:
Analysis: Degree distribution

Theorem 3: \( M(n, l, \mu, \Theta) \) follows a log-normal degree distribution as \( n \to \infty \) for some constant \( R \)

\[
\ln p_k \sim \mathcal{N} \left( \ln(n(\mu \beta + (1 - \mu)\gamma)^l) + l\mu \ln R + \frac{1}{2} l\mu(1 - \mu)(\ln R)^2, \quad l\mu(1 - \mu)(\ln R)^2 \right)
\]

if the network has a giant connected component.

- Simulation:
Theorem 4: \textit{MAG follows a power-law degree distribution}

\[ p_k \propto k^{-\delta - 0.5} \] for some \( \delta > 0 \)

when we set

\[ \frac{\mu_i}{1-\mu_i} = \left( \frac{\mu_i \alpha_i + (1-\mu_i) \beta_i}{\mu_i \beta_i + (1-\mu_i) \gamma_i} \right)^{-\delta} \]

Simulation:
Other network properties

![Graphs of network properties](image)

- **Yahoo!-Flickr**
- **MAG**

11/7/2012