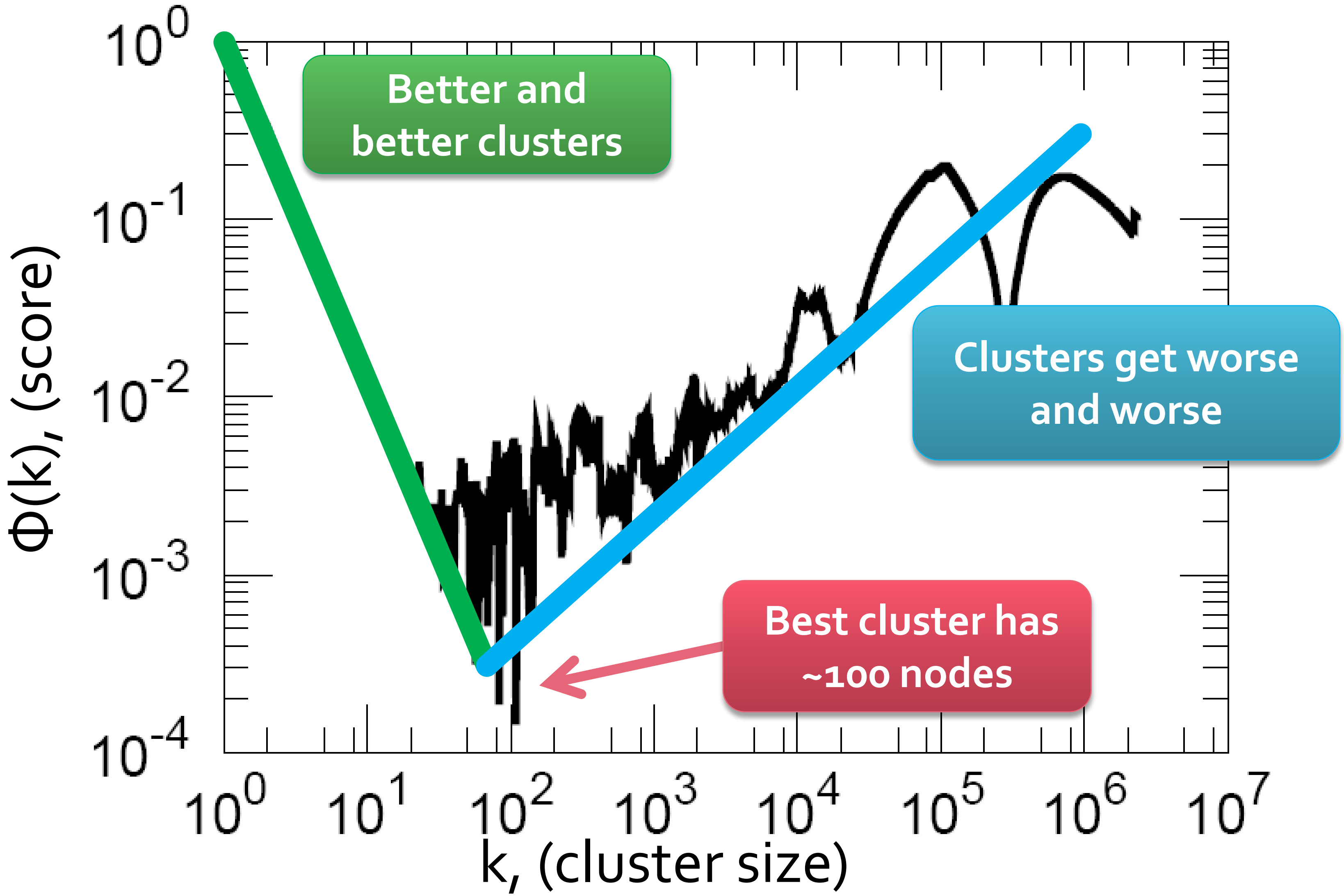
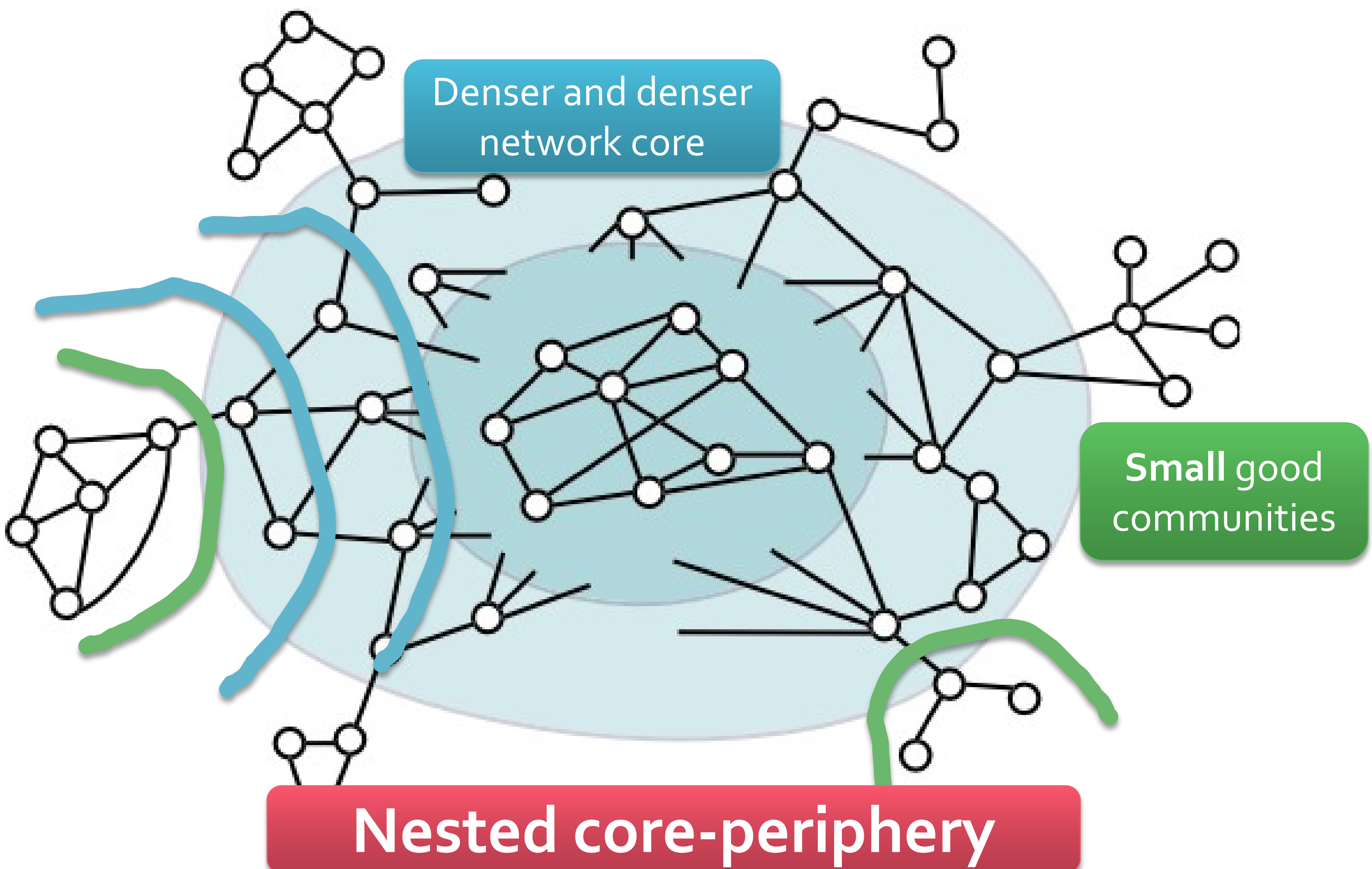
Kronecker graphs and the Structure of Large Networks

CS224W: Social and Information Network Analysis
Jure Leskovec, Stanford University
http://cs224w.stanford.edu

Recap: Network Community Profile

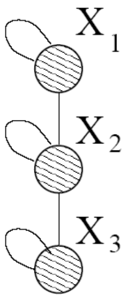


Explanation: Nested Core-Periphery

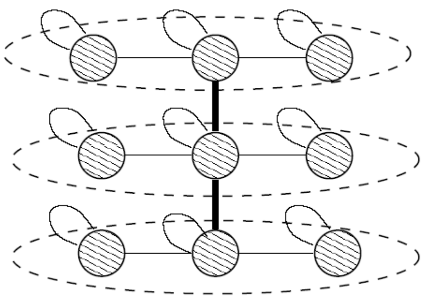


Idea: Recursive Graph Generation

- **Intuition:** Self-similarity
  - Object is similar to a part of itself (i.e. the whole has the same shape as one or more of the parts
- Mimic recursive graph / community growth



Initial graph

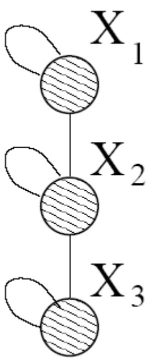


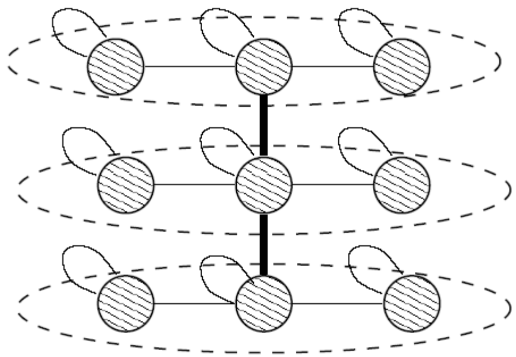
Recursive expansion

- **Kronecker Product** is a way of generating self-similar matrices

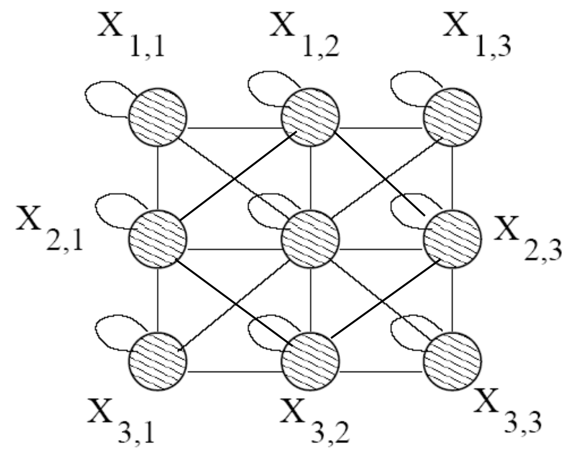
[PKDD ‘05]

Kronecker: Graph Growth





Intermediate stage



1	1	0
1	1	1
0	1	1

(3x3)

K_1

Initiator graph

K_1	K_1	0
K_1	K_1	K_1
0	K_1	K_1

(9x9)

$$K_2 = K_1 \otimes K_1$$

After the growth phase

Kronecker Product: Definition

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

$$\underset{N\,x\,M}{\mathbf{C} = \mathbf{A}} \otimes \underset{K\,x\,L}{\mathbf{B}} \doteq \begin{pmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \dots & a_{1,m}\mathbf{B} \\ a_{2,1}\mathbf{B} & a_{2,2}\mathbf{B} & \dots & a_{2,m}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1}\mathbf{B} & a_{n,2}\mathbf{B} & \dots & a_{n,m}\mathbf{B} \end{pmatrix}$$

$N*K\ x\ M*L$

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

Kronecker Product: Graph

- **Continuing multypling with K_1 we obtain K_4 and so on …**

1	1	0
1	1	1
0	1	1

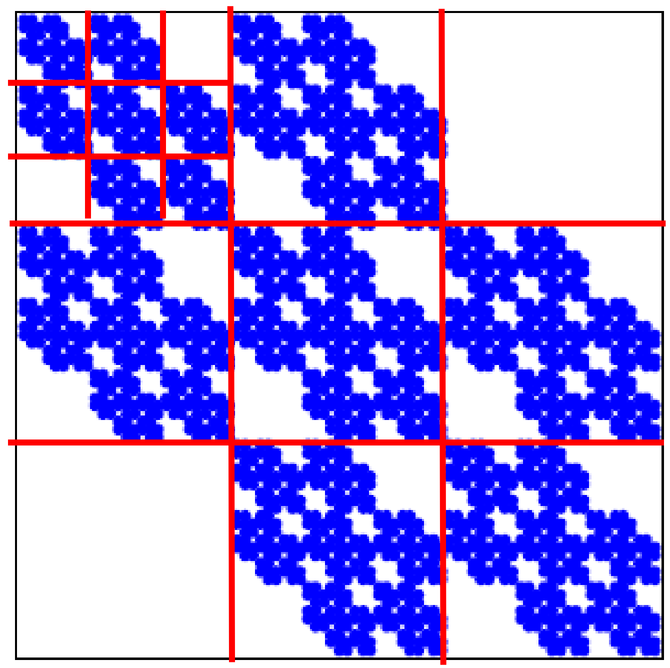
K_1

3 x 3

K_1	K_1	0
K_1	K_1	K_1
0	K_1	K_1

$$K_2 = K_1 \otimes K_1$$

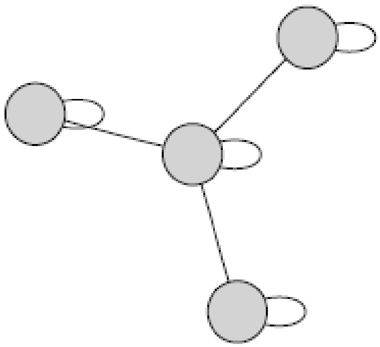
9 x 9



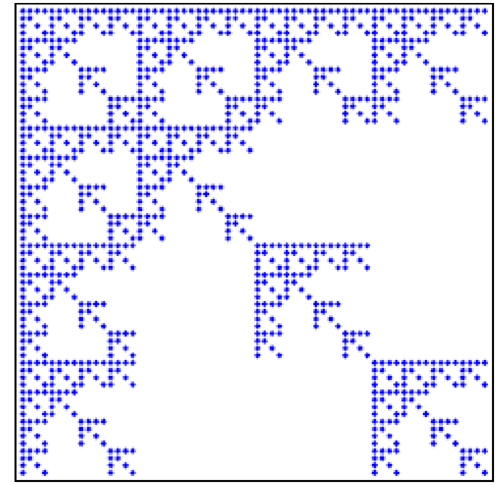
K_4 adjacency matrix

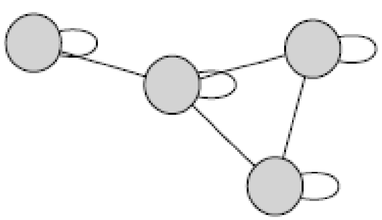
[PKDD '05]

Kronecker Initiator Matrices

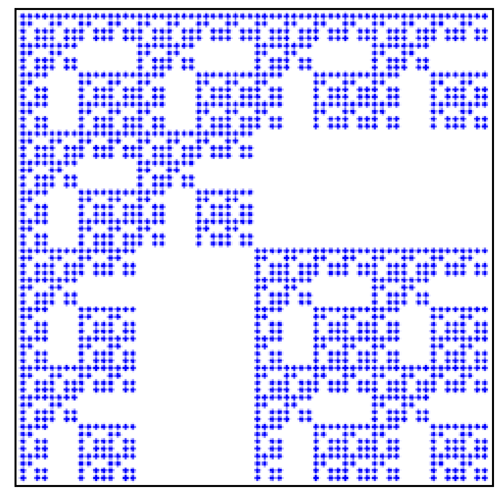


1	1	1	1
1	1	0	0
1	0	1	0
1	0	0	1





1	1	1	1
1	1	0	0
1	0	1	1
1	0	1	1

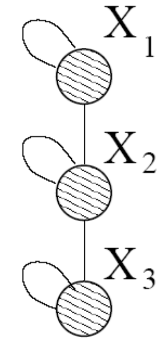


Initiator K_1 | K_1 adjacency matrix | K_3 adjacency matrix

Kronecker Graphs

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

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$



K_1

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

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

Kronecker Graphs

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

- 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
- We get **densification power-law:**
  - $\boldsymbol{E(t) \propto N(t)^a}$, What is a?
  - $\boldsymbol{a = \frac{\log(E(t))}{\log(N(t))} = \frac{\log(E_1)}{\log(N_1)}}$

1	1	0
1	1	1
0	1	1

K_1

Properties of Kronecker graphs

- **Kronecker graphs have many properties found in real networks:**
  - **Properties of static networks**
    - Power-Law like Degree Distribution
    - Power-Law eigenvalue and eigenvector distribution
    - Small Diameter
  - **Properties of dynamic networks**
    - Densification Power Law
    - Shrinking/Stabilizing Diameter

[PKDD '05]

Constant Diameter

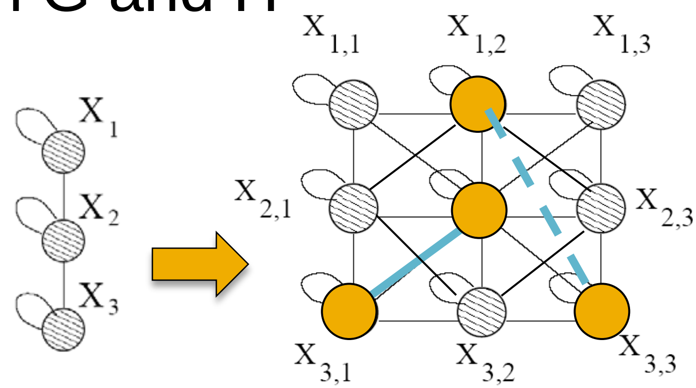
- Observation: Edges in Kronecker graphs:

$$Edge\ (X_{ij}, X_{kl}) \in G \otimes H$$

$$iff\ (X_i, X_k) \in G\ \ and\ (X_j, X_l) \in H$$

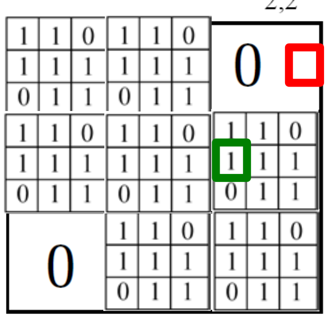
where X are appropriate nodes in G and H

- **Why?**
  - An entry in matrix G⊗H is a multiplication of entries in G and H.



Central node is $X_{2,2}$

1	1	0
1	1	1
0	1	1



Constant Diameter

- Theorem: **Constant diameter**: If G, H have diameter d then $G \otimes H$ has diameter d
- **What is distance between nodes u, v in $G \otimes H$?**
  - Let u=[a,b], v=[a',b'] (using notation from last slide) **then** edge (u,v) in $G \otimes H$ iif (a,a')$\in$G and (b,b')$\in$H
  - So, path a to a' in G is less *d* steps: $\boxed{a_1}, \boxed{a_2}, a_3, \ldots, a_d$
  - And path b to b' in H is less *d* steps: $\boxed{b_1}, \boxed{b_2}, b_3, \ldots, b_d$
  - **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 $G \otimes H$
- **Consequence:**
  - If K_1 has diameter d then graph K_k also has diameter d

[PKDD ’05]

Stochastic Kronecker Graphs

- Create $N_1 \times N_1$ probability matrix Θ_1
- Compute the k^{th} Kronecker power Θ_k
- For each entry p_{uv} of Θ_k include an edge (u,v) in K_k with probability p_{uv}

0.5	0.2
0.1	0.3

Θ_1

Kronecker multiplication

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

$\Theta_2 = \Theta_1 \otimes \Theta_1$

Probability of edge p_{ij}

Instance matrix K_2

flip biased coins

Stochastic Kronecker Graphs

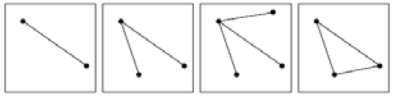
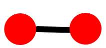
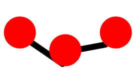
What is known about Stochastic Kronecker?

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

$$\Theta_1 = \begin{array}{|c|c|} \hline a & b \\ \hline b & c \\ \hline \end{array}$$

$a>b>c$

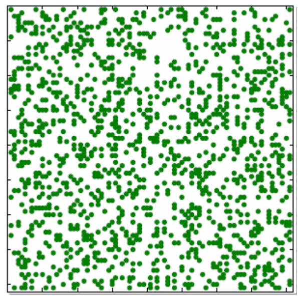
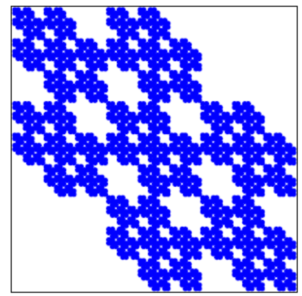
Estimating Kronecker graphs

- Given a real network G
  Want to estimate the initiator matrix: $\Theta_1 = \begin{array}{|c|c|} \hline a & b \\ \hline b & d \\ \hline \end{array}$
- **Method of moments** [Gleich&Owen '09]
  - Compare counts of  and solve the system of equations
  - For every of the 4 subgraphs, we get an equation:
    - 2 E[# ] = $(a+2b+c)^k$ - $(a+c)^k$ where $k = \log_2(N)$
    - 2 E[# ] = ...
    - ...
  - Now solve the system of equations by trying all possible values (a,b,c)

Kronecker graphs: Estimation

[ICML '07]

- **Maximum Likelihood Estimation**

$\arg\max_{\Theta_1} P($  $|$  $\xleftarrow{\text{Kronecker}} \Theta_1)$

- Naïve estimation takes *O(N!N²)*:
  - *N!* for different node labelings:
    - **Solution:** Metropolis sampling: *N!* → *(big) const*
  - N^2 for traversing graph adjacency matrix
    - **Solution:** Kronecker product ($E \ll N^2$): $N^2 \rightarrow E$
- Do gradient descent

$$\Theta_1 = \begin{array}{|c|c|} \hline a & b \\ \hline c & d \\ \hline \end{array}$$

Parameter Estimation: Approach

KronFit: Maximum likelihood estimation

- Given real graph G
- Find Kronecker initiator graph Θ (i.e., $\begin{array}{|c|c|}\hline a & b \\ \hline c & d \\ \hline\end{array}$) which

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

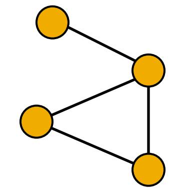
- We then need to (efficiently) calculate

$$P(G \mid \Theta)$$

- **And maximize over Θ**
  (e.g., using gradient descent)

KronFit: Likelihood $P(G|\Theta)$

- Given a graph G and Kronecker matrix Θ we calculate probability that Θ generated G $P(G|\Theta)$



G

0.5	0.2
0.1	0.3

Θ

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

Θ_k

1	0	1	1
0	1	0	1
1	0	1	1
1	1	1	1

G

$\boxed{P(G|\Theta)}$

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

Challenge 1: Node Correspondence

Θ

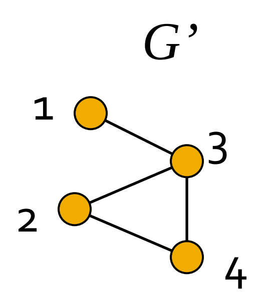
0.5	0.2
0.1	0.3

Θ_k

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

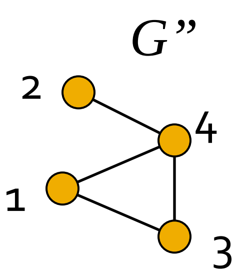
σ

G'



1	0	1	0
0	1	1	1
1	1	1	1
0	0	1	1

G"



1	0	1	1
0	1	0	1
1	0	1	1
1	1	1	1

$$P(G'|\Theta) = P(G''|\Theta)$$

- Nodes are unlabeled
- Graphs *G'* and *G"* should have the same probability *P(G'|Θ) = P(G"|Θ)*
- One needs to consider all node correspondences σ

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

- All correspondences are a priori equally likely
- There are ***O(N!)*** correspondences

Challenge 2: Calculating $P(G|\Theta,\sigma)$

- Assume that we solved the node correspondence problem
- Calculating

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

- Takes ***O(N²)*** time

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

Θ_k

σ

1	0	1	1
0	1	0	1
1	0	1	1
0	0	1	1

G

$$\boxed{P(G|\Theta,\sigma)}$$

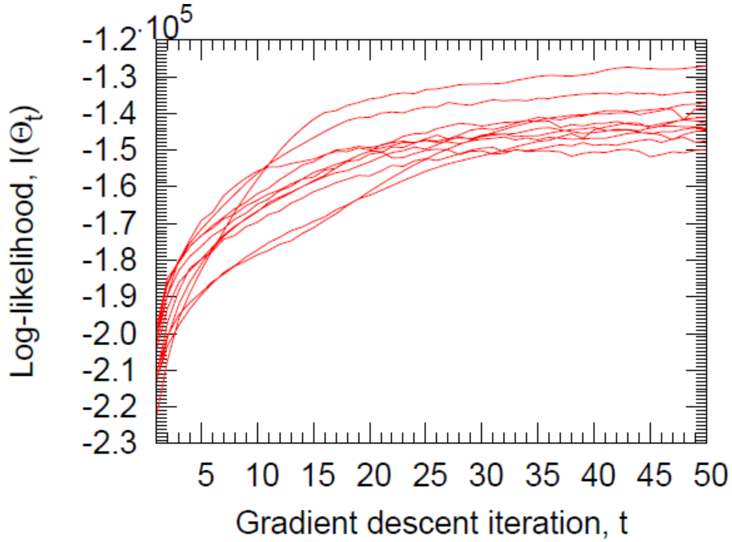
Experiments: real networks

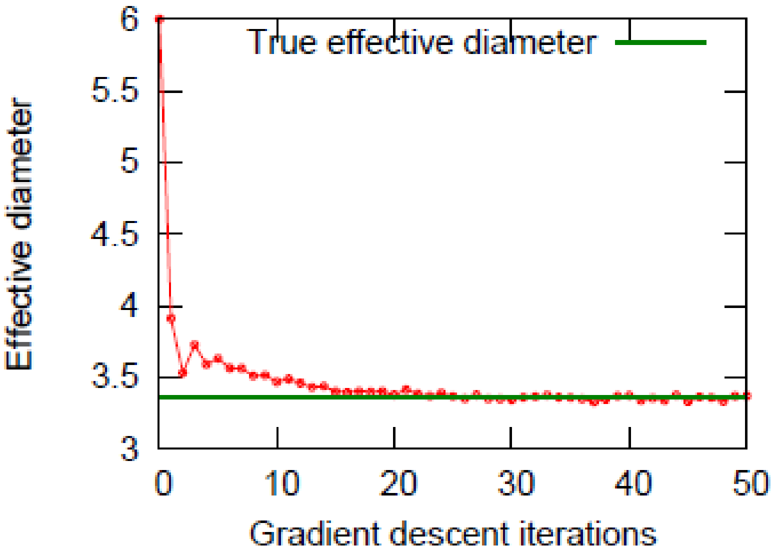
$$\Theta = \begin{array}{|c|c|} \hline a & b \\ \hline c & d \\ \hline \end{array}$$

- Experimental setup
  - Given real graph G
  - Gradient descent from random initial point
  - Obtain estimated parameters Θ
  - Generate synthetic graph K using Θ
  - 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

Convergence of fitting

- Can gradient descent recover true parameters?
  - Generate a graph from random parameters
  - Start at random point and use gradient descent
  - We recover true parameters 98% of the times

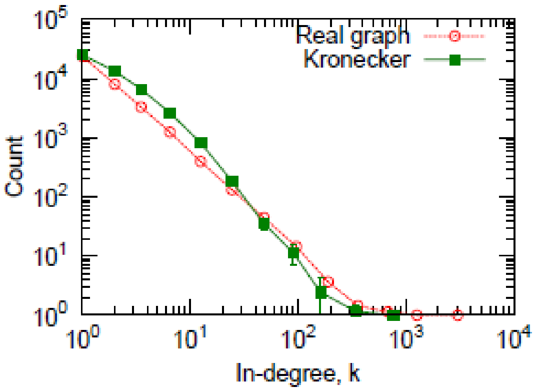




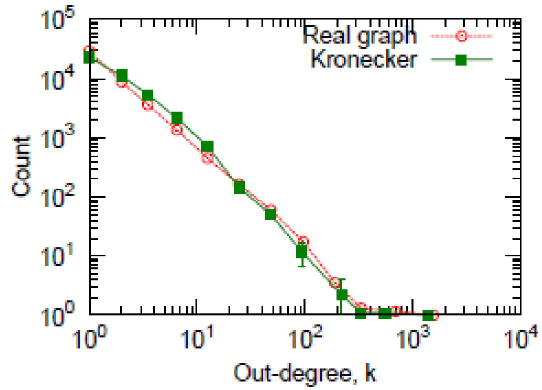
Estimation: Epinions (n=76k, m=510k)

- Real and Kronecker are very close:

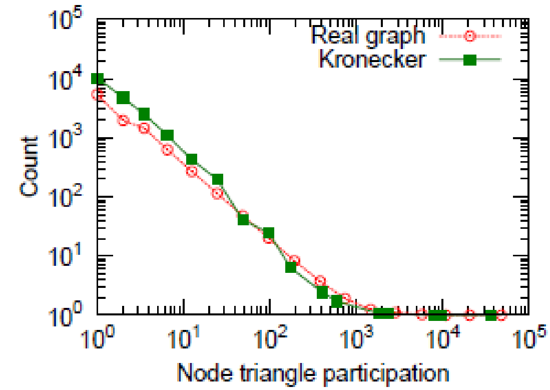
$$\Theta_1 = \begin{array}{|c|c|} \hline 0.99 & 0.54 \\ \hline 0.49 & 0.13 \\ \hline \end{array}$$



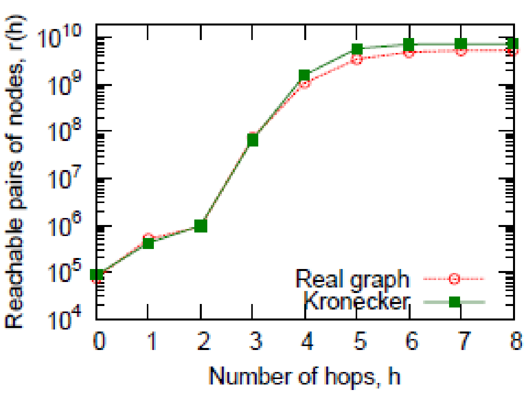
(a) In-Degree



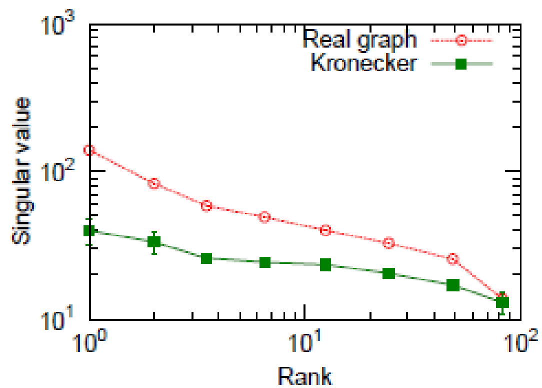
(b) Out-degree



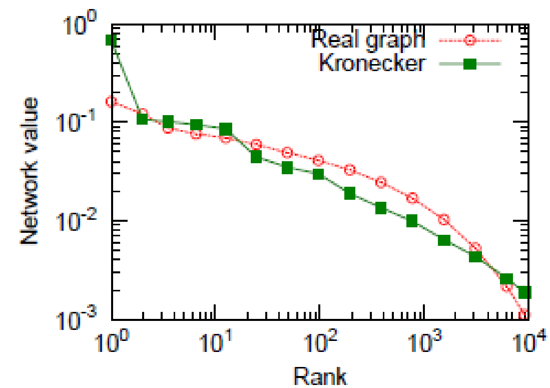
(c) Triangle participation



(d) Hop plot



(e) Scree plot



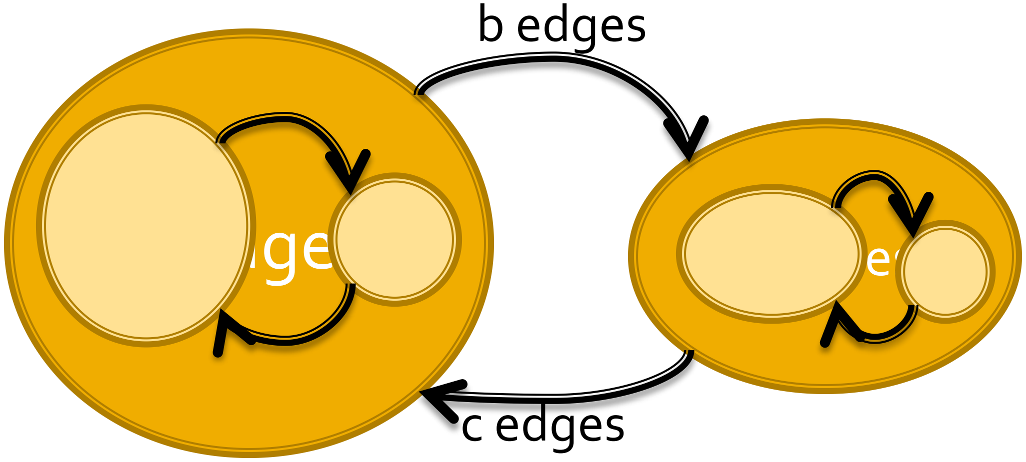
(f) "Network" value

[JMLR '10]

Kronecker & Network Structure

- **What do estimated parameters tell us about the network structure?**

$$\Theta = \begin{array}{|c|c|} \hline a & b \\ \hline c & d \\ \hline \end{array}$$

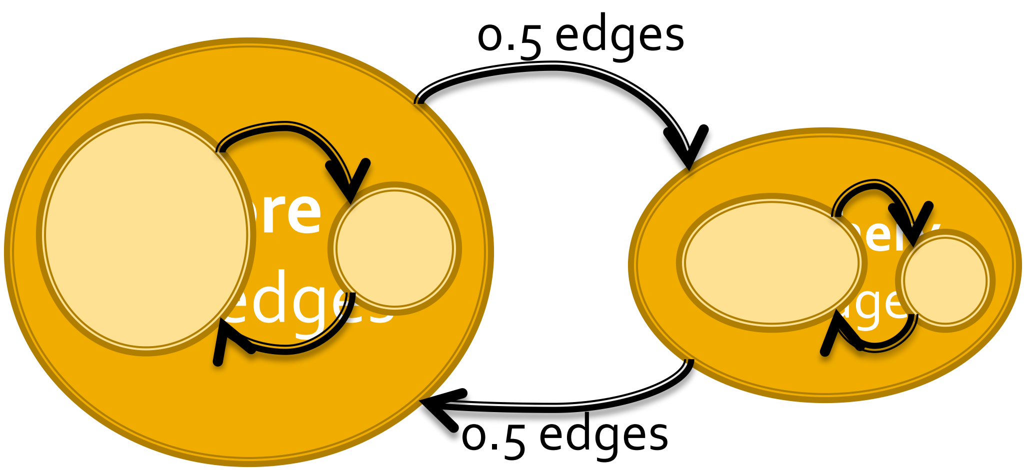


[JMLR '10]

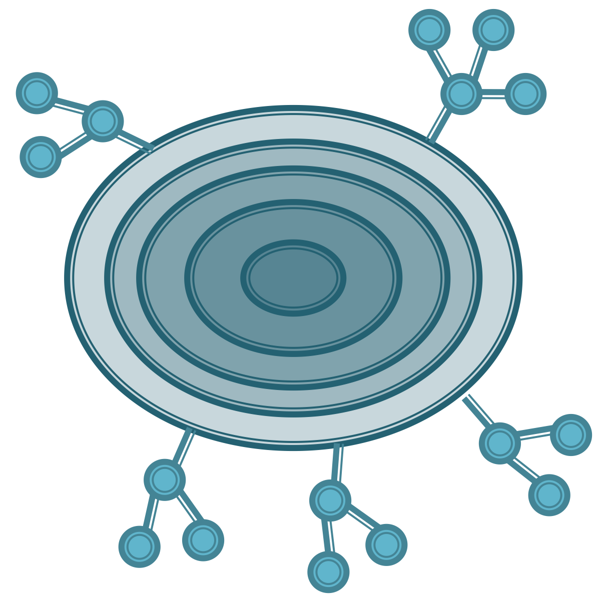
Kronecker & Network structure

- What do estimated parameters tell us about the network structure?

$$\Theta = \begin{array}{|c|c|}\hline 0.9 & 0.5 \\ \hline 0.5 & 0.1 \\ \hline\end{array}$$

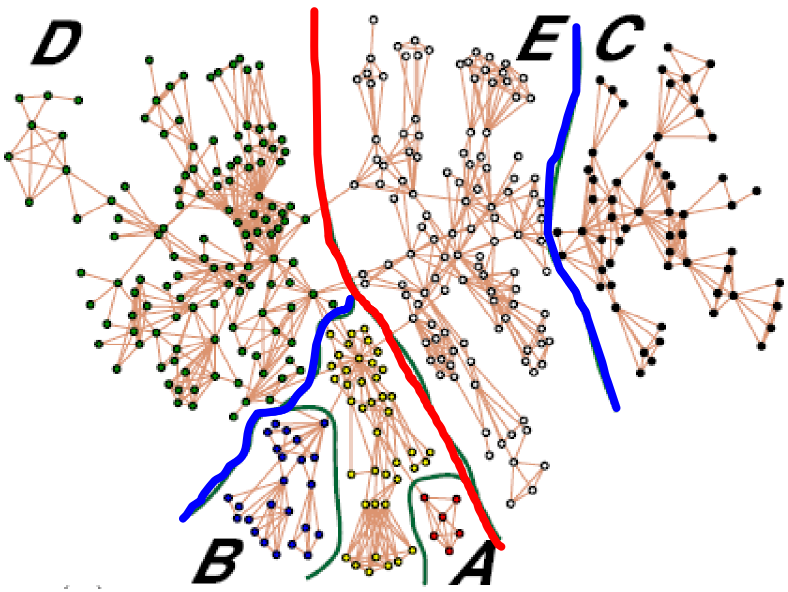


Nested Core-periphery



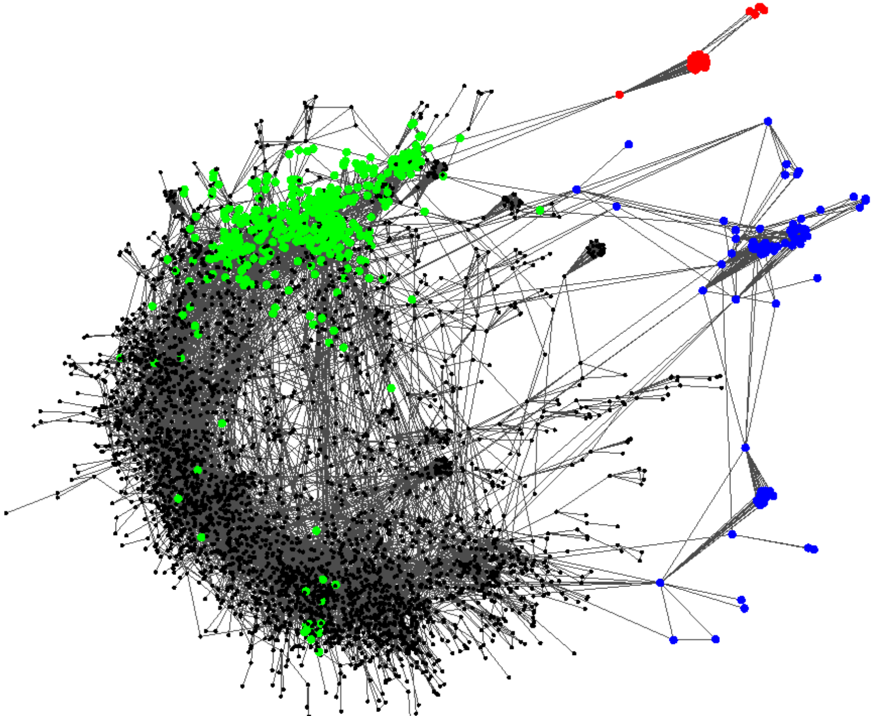
Small vs. Large Networks

- **Small and large networks are very different:**



$\Theta =$

0.99	0.17
0.17	0.82



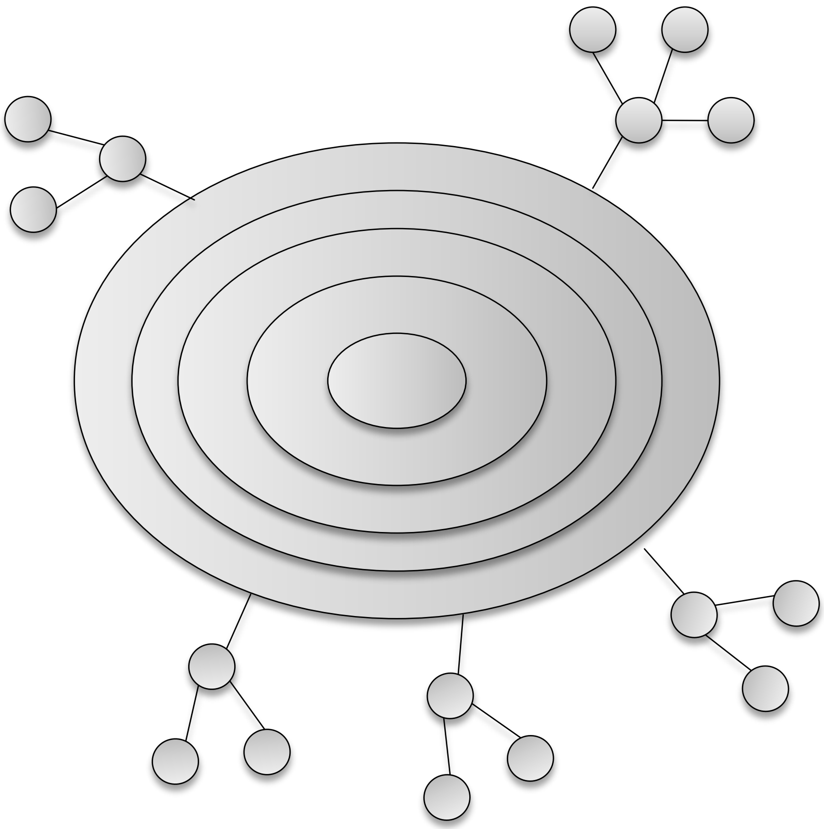
$\Theta =$

0.99	0.54
0.49	0.13

Implications (1)

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



Implications (2)

- 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)·(0.55+0.15) > 1. **Yes!**
- Real graphs are in the in the parameter region analogous to the giant component of an extremely sparse G_{np}

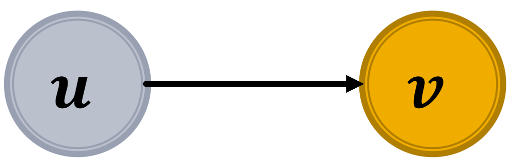
$$\Theta = \begin{bmatrix} 0.99 & 0.55 \\ 0.55 & 0.15 \end{bmatrix}$$

G_{np} — $1/n$ — real-networks — $\log(n)/n$

A Different Model: MAG Model

Nodes with Attributes

- 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?



u is friends with v

v's gender

u's gender (u \ v)	**FEMALE**	**MALE**
FEMALE	**0.3**	**0.6**
MALE	**0.6**	**0.2**

Link probability

Link-Affinity Matrix

- Let the values of the ***i-th attribute*** for node u and v be $\boldsymbol{a_i(u)}$ and $\boldsymbol{a_i(v)}$
  - $a_i(u)$ and $a_i(v)$ can take values $\{0, \cdots, d_i - 1\}$
- **Question:** How can we capture the influence of the attributes on link formation?
  - ***Attribute matrix $\boldsymbol{\Theta}$***

	$a_i(v) = 0$	$a_i(v) = 1$
$a_i(u) = 0$	$\boldsymbol{\Theta[0,0]}$	$\boldsymbol{\Theta[0,1]}$
$a_i(u) = 1$	$\boldsymbol{\Theta[1,0]}$	$\boldsymbol{\Theta[1,1]}$

$$\boxed{\boldsymbol{P(u,v) = \Theta[a_i(u), a_i(v)]}}$$

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

Approach: Great flexibility

- ***Flexibility* in the network structure:**
  - ***Homophily*** : love of the ***same***
    - e.g., political parties, hobbies

0.9	0.1
0.1	0.8

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

0.2	0.9
0.9	0.1

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

0.9	0.5
0.5	0.2

Combining attributes

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

$$a(u) = [\ 0 \quad 0 \quad 1 \quad 0\ ]$$
$$a(v) = [\ 0 \quad 1 \quad 1 \quad 0\ ]$$

$$\Theta_i = \begin{bmatrix} \alpha_1 & \beta_1 \\ \beta_1 & \gamma_1 \end{bmatrix} \begin{bmatrix} \alpha_2 & \beta_2 \\ \beta_2 & \gamma_2 \end{bmatrix} \begin{bmatrix} \alpha_3 & \beta_3 \\ \beta_3 & \gamma_3 \end{bmatrix} \begin{bmatrix} \alpha_4 & \beta_4 \\ \beta_4 & \gamma_4 \end{bmatrix}$$

$$P(u,v) = \alpha_1 \times \beta_2 \times \gamma_3 \times \alpha_4$$

Node attributes

+

Attribute matrices

↓

Link probability

Multiplicative Attribute Graph

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

$$\boldsymbol{P(u,v) = \prod_{i=1}^{l} \Theta_i[a_i(u), a_i(v)]}$$

Connection to Kronecker Graphs

- Initiator matrix K_1 acts like an **affinity** matrix
- **Probability of a link between nodes *u*, *v*:**

$$P(u,v) = \prod_{i=1}^{k} K_1(A_u(i), A_v(i))$$

$$K_1 = \begin{array}{cc|c} 0 & 1 & \\ \hline a & b & 0 \\ c & d & 1 \end{array}$$

	v_1	v_2	v_3	v_4
v_1	a·a	a·b	b·a	b·b
v_2	a·c	a·d	b·c	b·d
v_3	c·a	c·b	d·a	d·b
v_4	c·c	c·d	d·c	d·d

$K_2 = K_1 \otimes K_1$

=

	v_1	v_2	v_3	v_4
v_1	a	b	a	b
	a		**b**	
v_2	c	d	c	d
v_3	a	b	a	b
	c		**d**	
v_4	c	d	c	d

$v_2 = (0,1)$

$v_3 = (1,0)$ $\quad P(v_2,v_3) = b \cdot c$

Connection to Kronecker Graphs

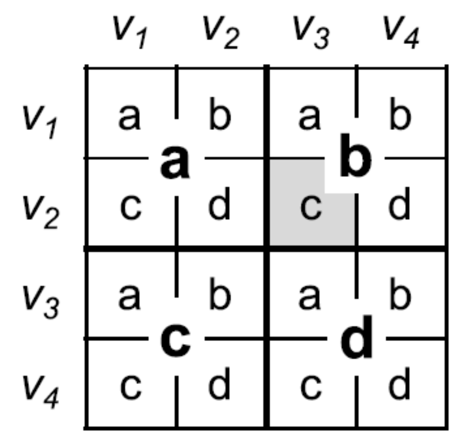
- Each node in a Kronecker graph has a node id (e.g. $0, \cdots, 2^l - 1$)
- A binary representation of node id is its attribute vector in a MAG model
- Then, the (stochastic) adjacency matrices of two models are equivalent
- **Example:**

$$\boldsymbol{K} \quad \begin{array}{|c|c|} \hline a & b \\ \hline c & d \\ \hline \end{array} \quad \Rightarrow \quad \begin{array}{c|cccc|} & v_0 & v_1 & v_2 & v_3 \\ \hline v_0 & a & b & a & b \\ v_1 & c & d & \boxed{c} & d \\ v_2 & a & b & a & b \\ v_3 & c & d & c & d \\ \hline \end{array}$$

$$\boxed{\begin{aligned} a(v_1) &= [0 \;\; 1] \\ a(v_2) &= [1 \;\; 0] \\ P(v_1, v_2) &= b \cdot c \end{aligned}}$$

Feature vector view: Question

- **2 ingredients of Kronecker model:**
  - (1) Each of 2^k nodes has a unique binary vector of length k
    - Node id expressed binary is the vector
  - (2) The initiator matrix K



- **Question:**
  - What if ingredient (1) is dropped?
    - i.e., do we need high variability of feature vectors?

Comparison: Adjacency matrices

- **Adjacency matrices:**