Networks: Modeling Interactions

Jure Leskovec (@jure)
Stanford University

KDD Summer School, Beijing, August 2012

Models for networks

- **Graph:**
  Kronecker graphs
- **Graph + Node attributes:**
  MAG model
- **Graph + Edge attributes:**
  Signed networks
- **Link Prediction/Recommendation:**
  Supervised Random Walks

Networks with Metadata

- **Many networks come with:**
  - The graph (wiring diagram)
  - Node/edge metadata (attributes/features)
- **How to generate realistic looking graphs?**
  - **1: Kronecker Graphs**
- **How to model networks with node attributes?**
  - **2: Multiplicative Attributes Graph (MAG) model**
- **How to model networks with edge attributes?**
  - **3: Networks of Positive and Negative Edges**
- **How to predict/recommend new edges?**
  - **4: Supervised Random Walks**

Want to learn more? (1)

- **Stanford Large Network Dataset Collection**
  - http://snap.stanford.edu
  - 60+ large networks:
    - Social network, Geo-location networks, Information networks, Evolving networks, Citation networks, Internet networks, Amazon, Twitter, …
- **Stanford Network Analysis Platform (SNAP):**
  - http://snap.stanford.edu
  - C++ Library for massive networks
  - Has no problem working with 1B nodes, 10B edges

Want to learn more? (2)

- **Stanford CS224W: Social and Information Networks Analysis**
  - http://cs224w.stanford.edu
  - Graduate course on topics discusses today
  - Slides, homeworks, readings, data, …
- **My webpage**
  - http://cs.stanford.edu/~jure/
  - Videos of talks and tutorials
- **Twitter: @jure**

Kronecker Graphs Model

Reliably models the global network structure using only 4 parameters!

The Setting

- Want to have a model that can generate a realistic networks with realistic growth:
  - **Static Patterns**
    - Power Law Degree Distribution
    - Small Diameter
    - Power Law Eigenvalue and Eigenvector Distribution
  - **Temporal Patterns**
    - Densification Power Law
    - Shrinking/Constant Diameter
- **For Kronecker graphs:**

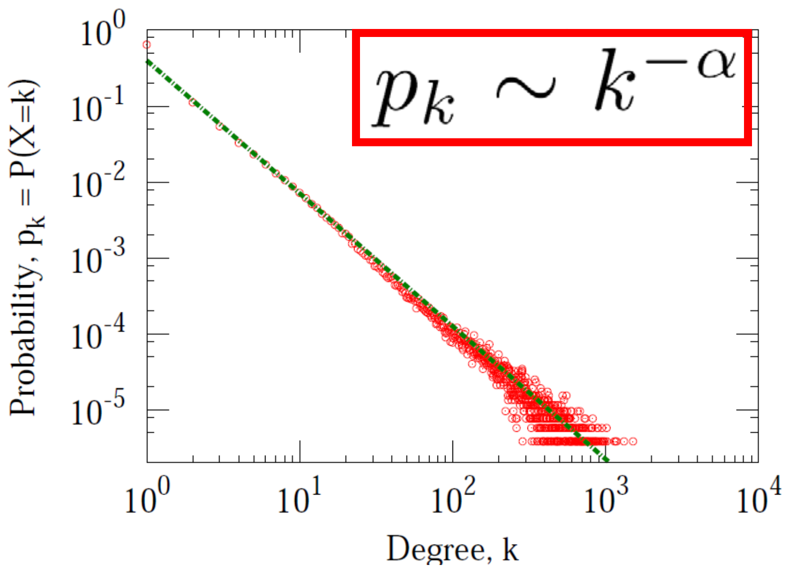
  1) analytically tractable (i.e, prove power-laws, etc.)

  2) statistically interesting (i.e, fit it to real data)

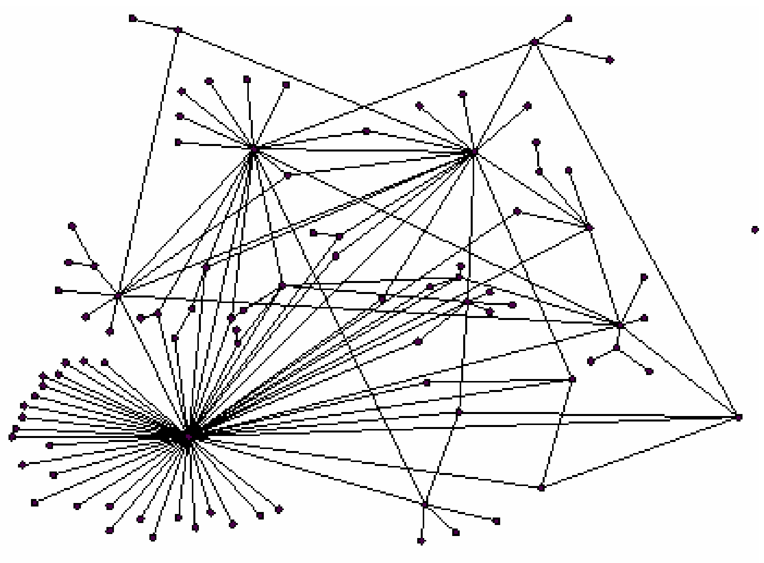
Degree Distribution

- **Classical example:**

  Heavy-tailed degree distributions



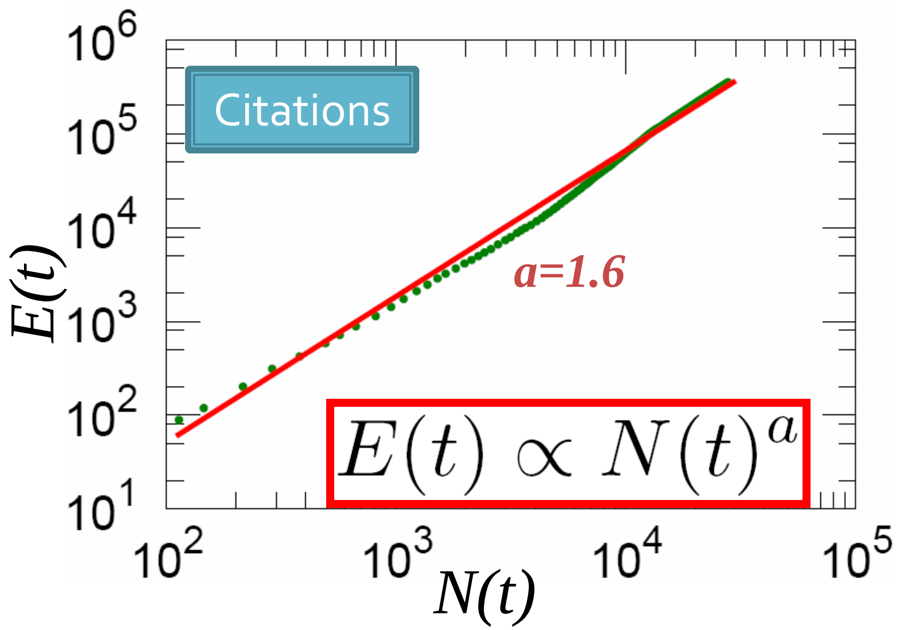
Flickr social network
n= 584,207, m=3,555,115

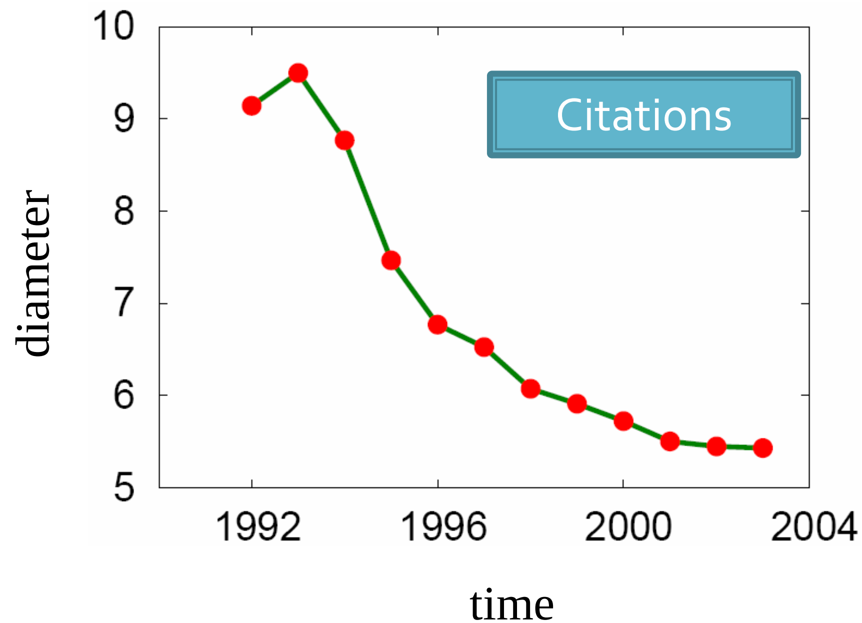


Scale free networks
many hub nodes

Scaling of Network Properties

- How do network properties scale with the size of the network?





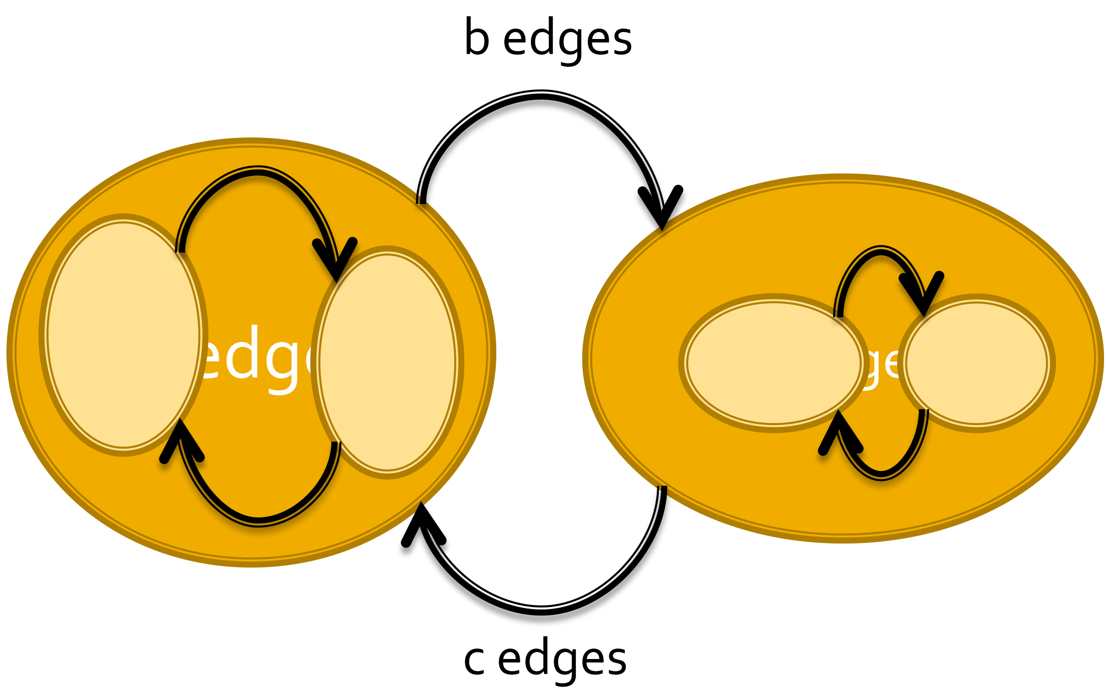
Densification
Average degree increases

Shrinking diameter
Path lengths get shorter

Recursive Model of Networks

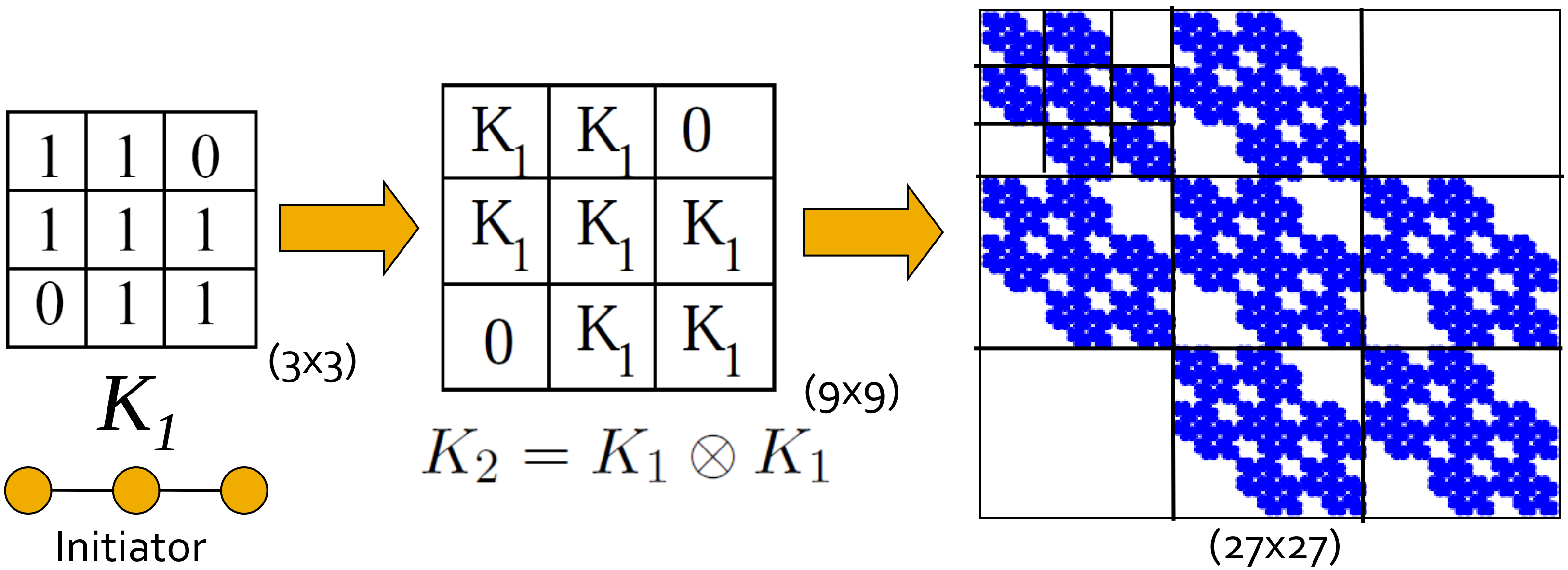
- **How can we think of network structure recursively?**

$$K_1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$



Recursive model of network

- **Kronecker graphs:**
  - **A recursive model of network structure**



Kronecker Graphs

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

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

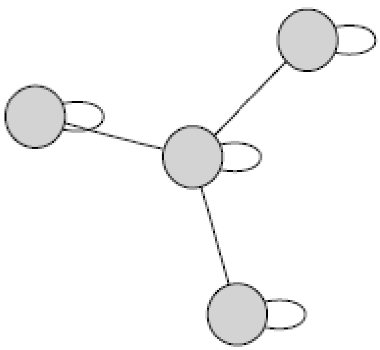
$N*K \; x \; M*L$

- Define: Kronecker product of two graphs is a Kronecker product of their adjacency matrices
- **Kronecker graph:** a growing sequence of graphs by iterating the Kronecker product

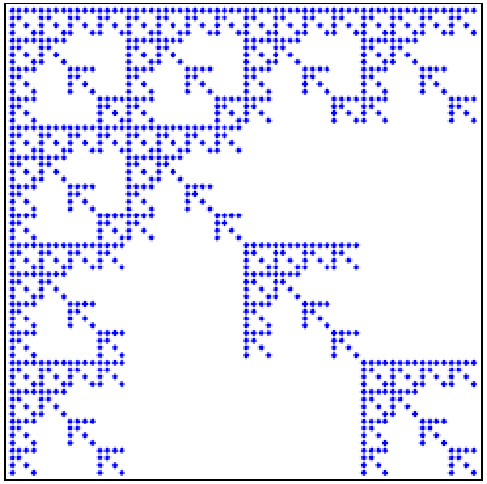
$$K_1^{[k]} = K_k = \underbrace{K_1 \otimes K_1 \otimes \dots K_1}_{k \; times} = K_{k-1} \otimes K_1$$

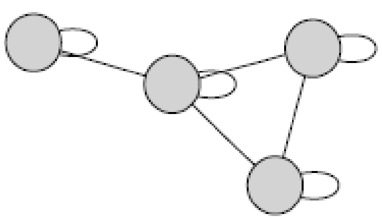
[Leskovec et al. PKDD '05]

Kronecker Initiator Matrices

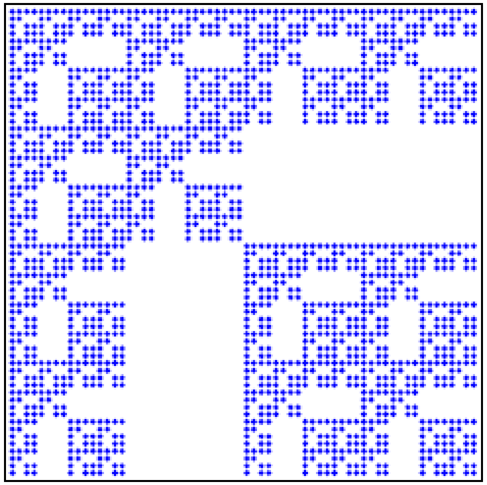


1	1	1	1
1	1	0	0
1	0	1	0
1	0	0	1





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

Properties Kronecker Graphs

- **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
  - Properties of evolving networks:
    - Densification Power Law
    - Shrinking/Stabilizing Diameter
- **Can we make the model stochastic?**

Stochastic Kronecker Graphs

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

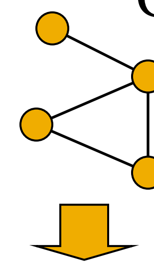
Instance matrix K_2

flip biased coins

Kronecker: Parameter Estimation

- **Given** a graph G
- **What is the parameter matrix Θ?**
- Find Θ that maximizes P(G|Θ)

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

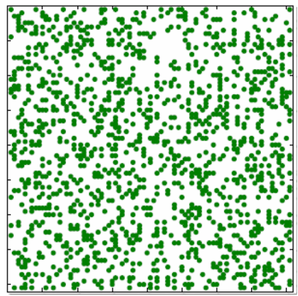
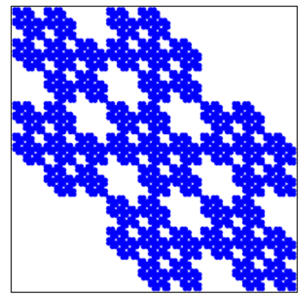
$\mathrm{P}(G|\Theta)$

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

[Leskovec-Faloutsos ICML '07]

Kronecker: Parameter Estimation

- **Maximum likelihood estimation**

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

$\Theta_1 =$

a	b
c	d

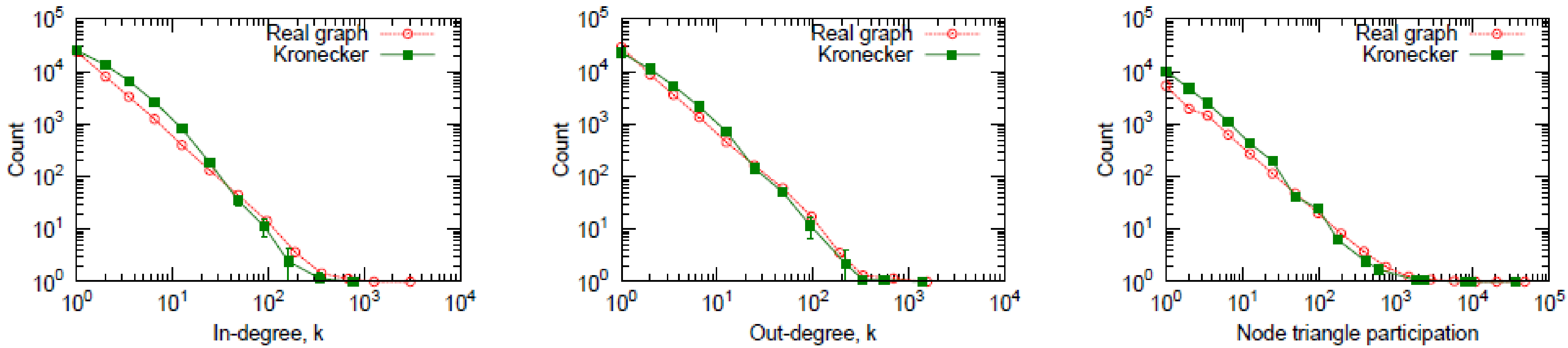
- Naïve estimation takes $O(N!N^2)$:
  - $N!$ for different node labelings:
  - N^2 for traversing graph adjacency matrix
- Do gradient descent

We estimate the model in O(E)

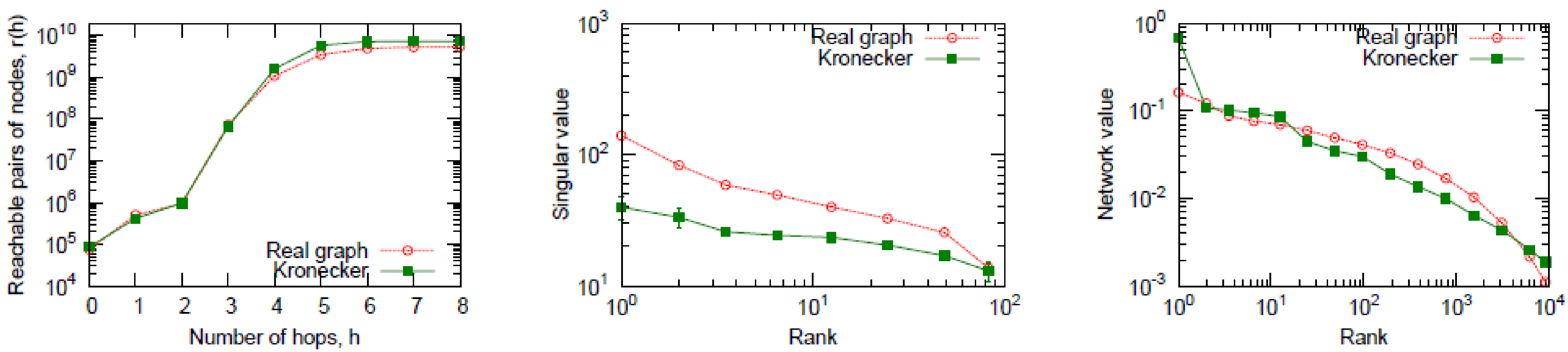
Epinions (n=76k, m=510k)

- **Real and Kronecker are very close:**

$$\Theta_1 = \begin{bmatrix} 0.99 & 0.54 \\ 0.49 & 0.13 \end{bmatrix}$$



(a) In-Degree (b) Out-degree (c) Triangle participation



(d) Hop plot (e) Scree plot (f) "Network" value

The MAG Model

- For networks with node attributes
- Can do power-law and log-normal degrees

Modeling Questions

- **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:***

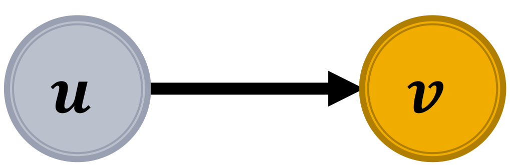
    1) Analytically tractable (prove network properties)

    2) Statistically interesting (can be fit to real data)

[Internet Math. '12]

Our Approach: Node attributes

- **Each node has a set of categorical attributes**
  - Gender: Male, Female
  - Home country: US, Canada, Russia, etc.
- **How do node attributes influence link formation?**
  - **Example:** MSN Instant Messenger [Leskovec&Horvitz '08]



Chatting network

v's gender

u's gender: u \ v	FEMALE	MALE
FEMALE	0.3	0.7
MALE	0.7	0.3

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?**
  - Key: ***Attribute link-affinity matrix*** $\boldsymbol{\Theta}$

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

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

  - Each entry captures the ***affinity of a link*** between two nodes associated with the attributes of them

Link-Affinity Matrix

- ***Link-Afinity Matrices*** **offer *flexibility in modeling the network structure:***
  - ***Homophily*** : love of the ***same***
    - e.g., political views, 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

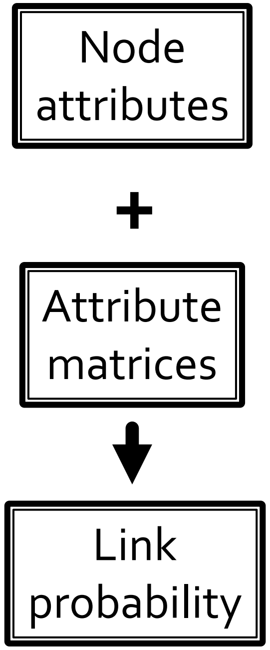
From Attributes to Links

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

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

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

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



[Internet Math. '12]

Multiplicative Attribute Graph

- **MAG model** $M(n, l, A, \vec{\Theta})$ **:**
  - A network contains $\boldsymbol{n}$ **nodes**
  - Each node has $\boldsymbol{l}$ **categorical attributes**
  - $A = [a_i(u)]$ represents the $\boldsymbol{i}$**-th attribute of node** $\boldsymbol{u}$
  - Each attribute can take $\boldsymbol{d_i}$ **different values**
  - Each attribute has a $\boldsymbol{d_i} \times \boldsymbol{d_i}$ **link-affinity matrix** $\boldsymbol{\Theta_i}$
  - Edge probability between nodes u and v

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

Analysis: MAG Model

- **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***

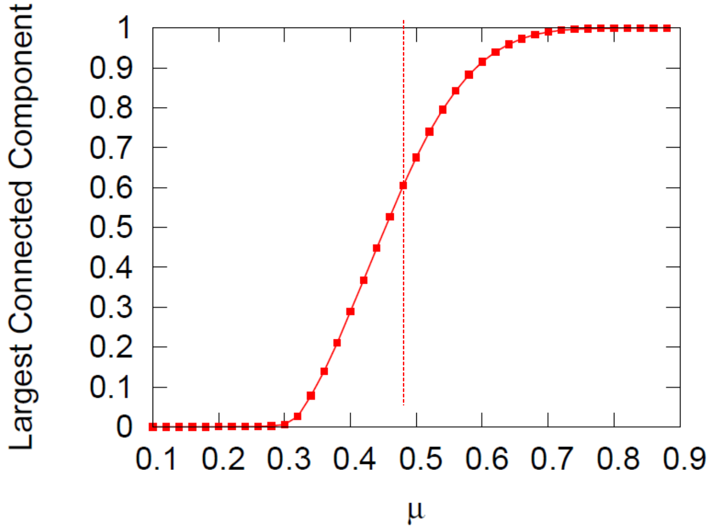
Analysis: Connected component

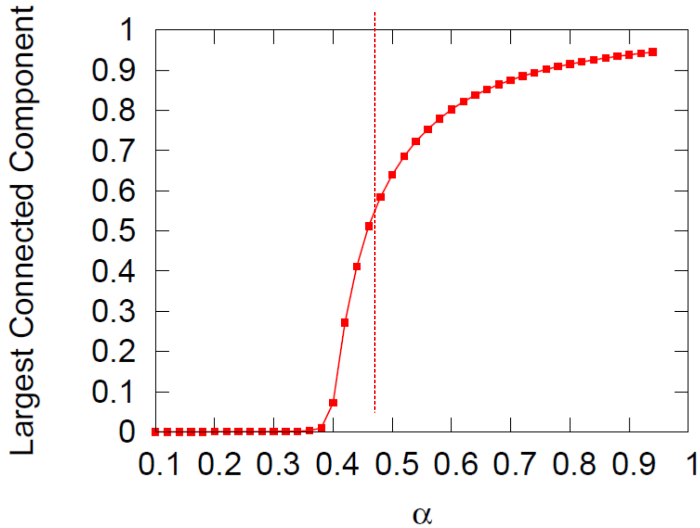
Theorem 1: ***A unique giant connected component*** *of size $\theta(n)$ exists in $M(n, l, \mu, \Theta)$ w.h.p. as $n \to \infty$ if*

$$P(a_i(u) = 1) = \mu$$

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

- **Simulation:**





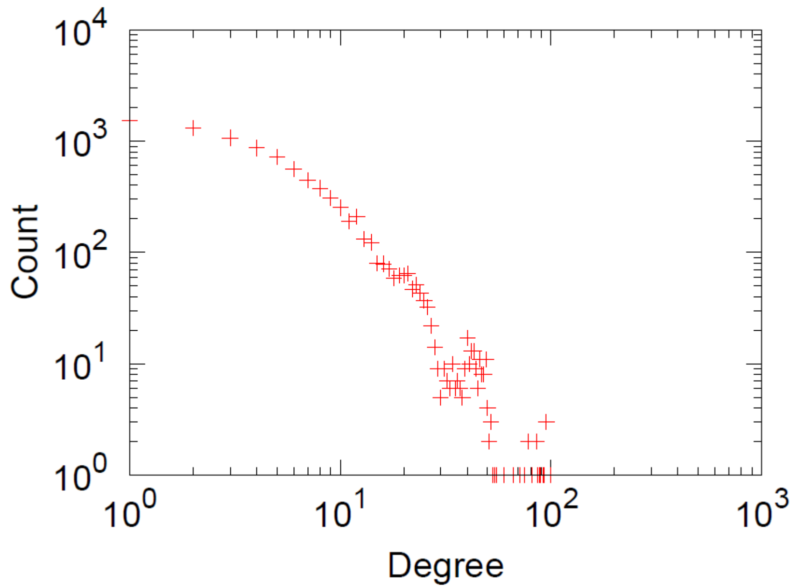
Analysis: Degree distribution

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

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

if the network has a giant connected component.

- **Simulation:**



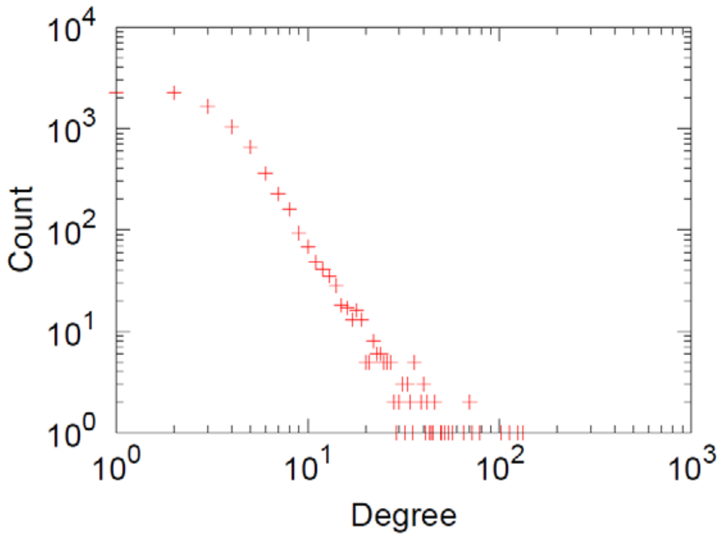
Analysis: Power-law in MAG

Theorem 4: *MAG follows* ***a power-law degree distribution***

$$p_k \propto k^{-\delta-0.5} \quad \text{for some } \delta > 0$$

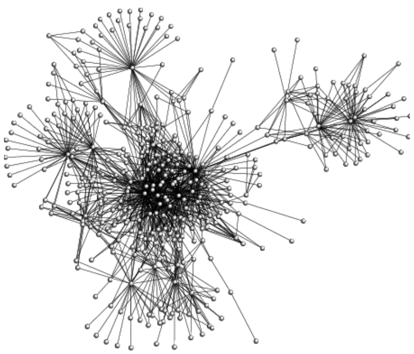
$$\text{when we set } \frac{\mu_i}{1-\mu_i} = \left(\frac{\mu_i\alpha_i+(1-\mu_i)\beta_i}{\mu_i\beta_i+(1-\mu_i)\gamma_i}\right)^{-\delta}$$

- **Simulation:**



Fitting the MAG model

- **MAG model is also statistically "interesting"**
- Estimate model parameters from the data
  - **Given:**
    Links of the network
  - **Estimate:**
    - Node attributes
    - Link-affinity matrices
- Formulate as a maximum likelihood problem
- **Solve it using variational EM**



$$a(u) = [\dots]$$

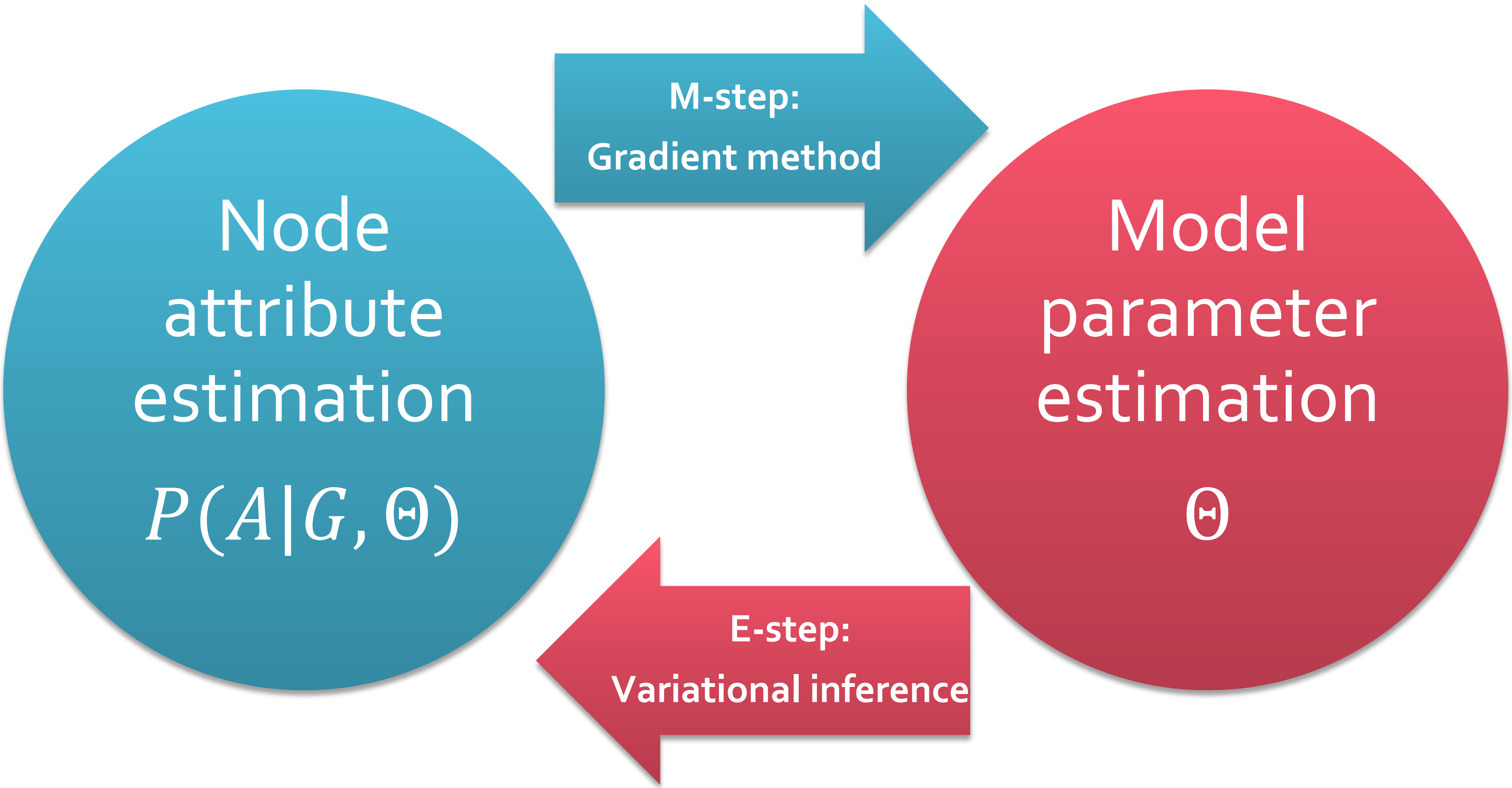
$$\Theta_i =$$

0.9	0.1
0.1	0.8

Fitting the MAG model

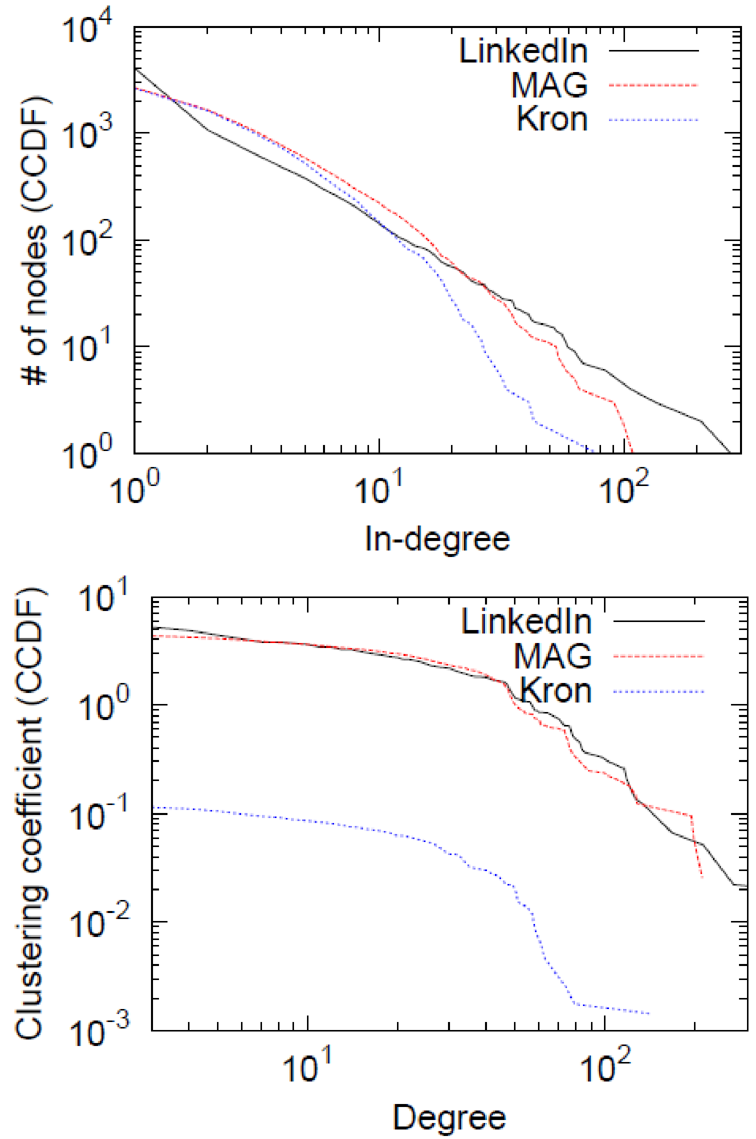
- **Edge probability:**
  - $\boldsymbol{P(u,v) = \prod_{i=1}^{l} \Theta_i[a_i(u), a_i(v)]}$
- **Network likelihood:**
  - $P(G|A,\Theta) = \prod_{G_{uv}=1} P(u,v) \cdot \prod_{G_{uv}=0} 1 - P(u,v)$
    - G … graph adjacency matrix
    - A … matrix of node attributes
    - Θ… link-affinity matrices
- **Want to solve:**
  - $\arg\max_{A,\Theta} P(G|A,\Theta)$

Variational EM



Experiments: Global Structure

- **LinkedIn network**
  - When it was super-young (4k nodes, 10k edges)
- Fit using 11 latent binary attributes per node



Experiments: AddHealth

- Case study: **AddHealth**
  - School friendship network
  - Largest network: 457 nodes, 2259 edges
  - Over 70 school-related attributes for each student
  - Real features are selected in the greedy way to maximize the likelihood of MAG model
    - We fit only Θ (since A is given): $\arg\max_{\Theta} P(G, A|\Theta)$
    - 7 features
- **Model accurately fits the network structure**

Experiments: AddHealth

- **Most important features for tie creation**

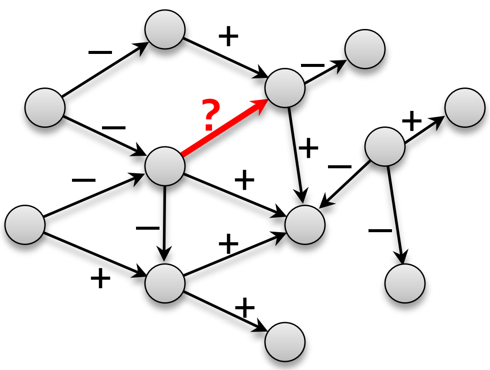
Affinity matrix	Attribute description
[0.572 0.146; 0.146 0.999]	School year (0 if $\geq$ 2)
[0.845 0.332; 0.332 0.816]	Highest level math (0 if $\geq$ 6)
[0.788 0.377; 0.377 0.784]	Cumulative GPA (0 if $\geq$ 2.65)
[0.999 0.246; 0.246 0.352]	AP/IB English (0 if taken)
[0.794 0.407; 0.407 0.717]	Foreign language (0 if taken)

Models of Networks with Signed Edges

- How people determine friends and foes?
- Predict friend vs. foe with 90% accuracy

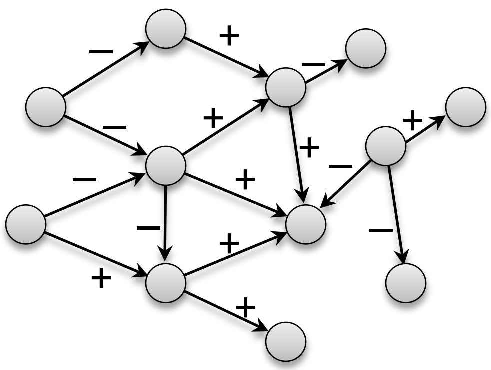
Friends vs. Foes

- **So far we viewed links as positive but links can also be negative**
- **Question:**
  - How do edge signs and network interact?
  - How to model and predict edge signs?
- **Applications:**
  - Friend recommendation
    - Not just whether you know someone but what do you think of them



Networks with Explicit Signs

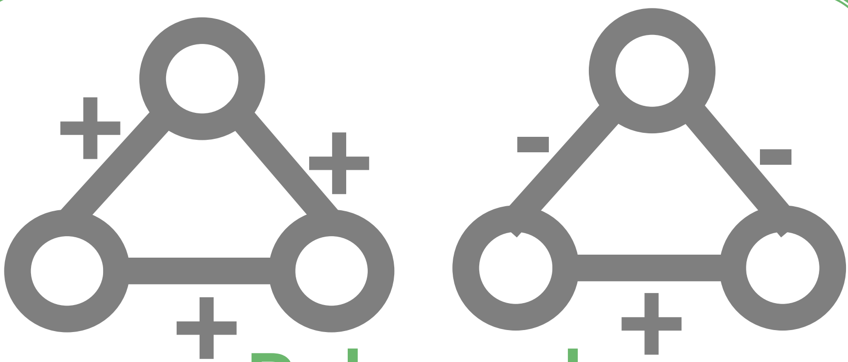
- Each link A→B is **explicitly** tagged with a sign:
  - Epinions: Trust/Distrust
    - Does A trust B's product reviews?
      (only positive links are visible)
  - Wikipedia: Support/Oppose
    - Does A support B to become Wikipedia administrator?
  - Slashdot: Friend/Foe
    - Does A like B's comments?
  - Other examples:
    - Sentiment analysis of the communication



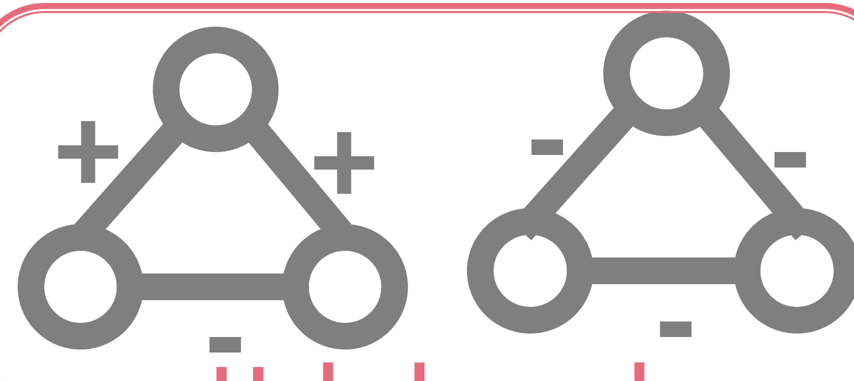
	Epinions	Slashdot	Wikipedia
Nodes	119,217	82,144	7,118
Edges	841,200	549,202	103,747
+ edges	85.0%	77.4%	78.7%
– edges	15.0%	22.6%	21.2%

Theory of Structural Balance

- **Start with intuition** [Heider '46]:
  - Friend of my friend is my friend
  - Enemy of enemy is my friend
  - Enemy of friend is my enemy
- Look at connected triples of nodes:



Balanced



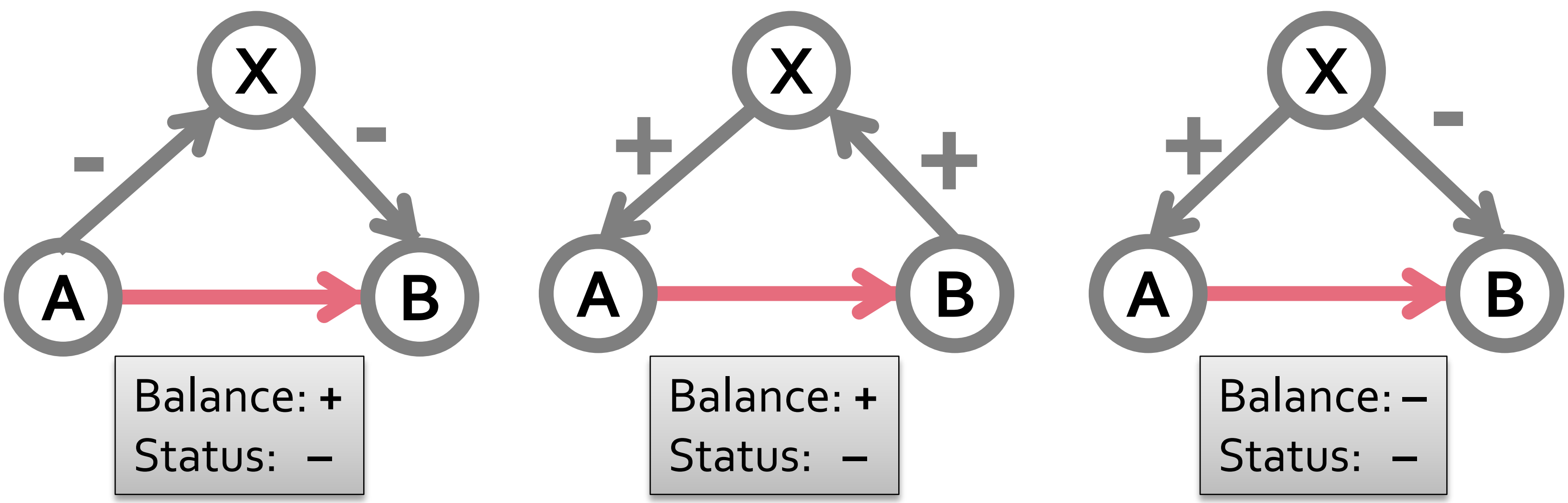
Unbalanced

Consistent with "friend of a friend" or "enemy of the enemy" intuition

Inconsistent with the "friend of a friend" or "enemy of the enemy" intuition

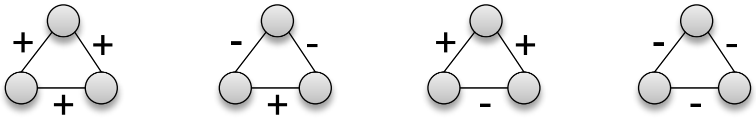
Theory of Status

- **Status theory** [Davis-Leinhardt '68, Leskovec et al. '10]
  - Link $A \xrightarrow{+} B$ means: B has higher status than A
  - Link $A \xrightarrow{-} B$ means: B has lower status than A
  - Signs/directions of links to X make a prediction
- **Status and balance make different predictions:**

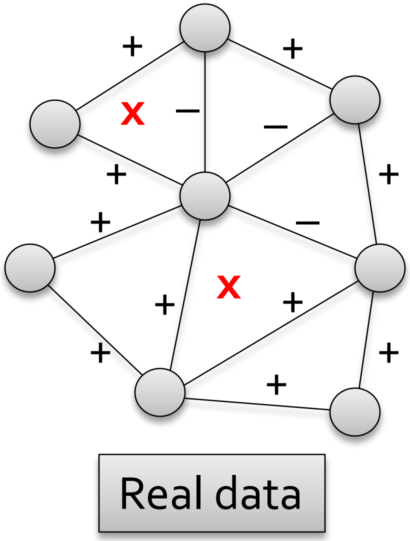


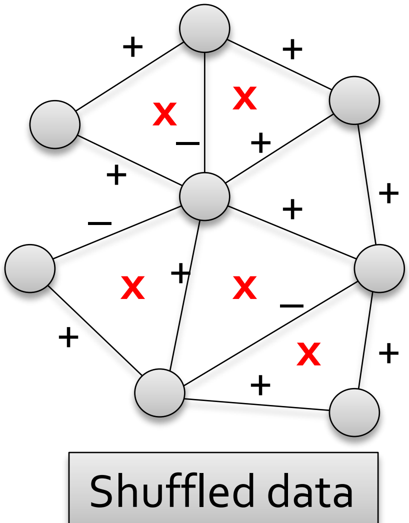
Undirected Links: Balance

- Consider networks as undirected
- Compare frequencies of signed triads in real and shuffled data
  - 4 triad types *t*:



- **Surprise** value for triad type *t*:
  - Number of std. deviations by which number of occurrences of triad *t* differs from the expected number in shuffled data





Undirected Links: Balance

- Surprise values: *i.e.,* z-score (deviation from random measured in the number of std. devs.)

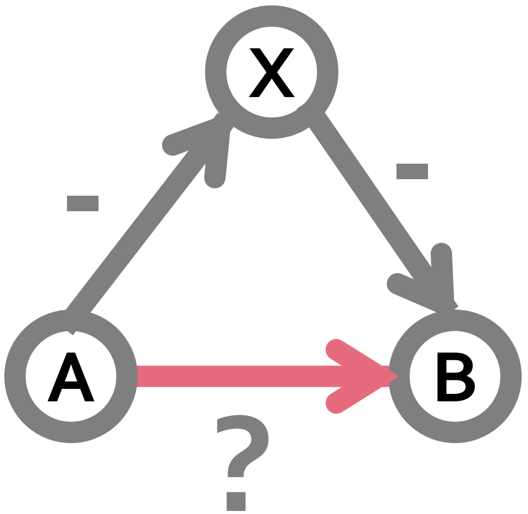
	Triad	Epin	Wiki	Slashdot
Balanced	+ + +	1,881	380	927
Balanced	- - +	249	289	-175
Unbalanced	+ + -	-2,105	-573	-824
Unbalanced	- - -	288	11	-9

- Observations:
  - Strong signal for balance
  - Epinions and Wikipedia agree on all types
  - Consistency with Davis's ['67] weak balance

Evolving Directed Networks

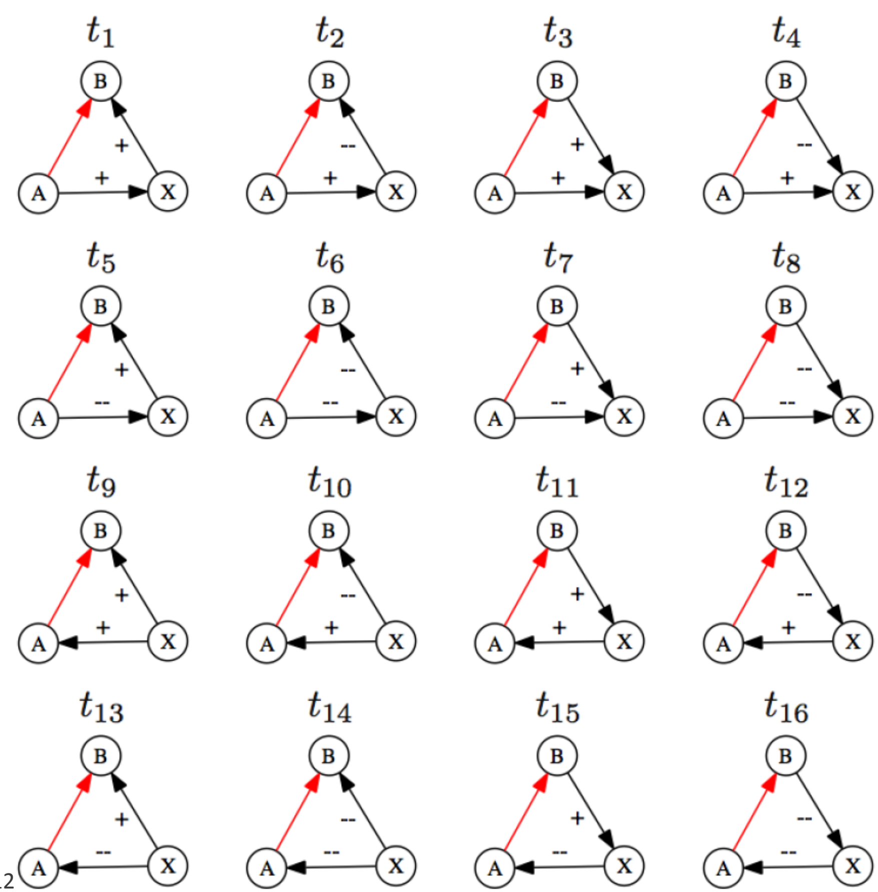
[CHI '10]

- Links are directed and created over time
- To compare balance and status we need to formalize two issues:
  - Links are embedded in triads which provide **contexts for signs**
  - Users are **heterogeneous** in their linking behavior



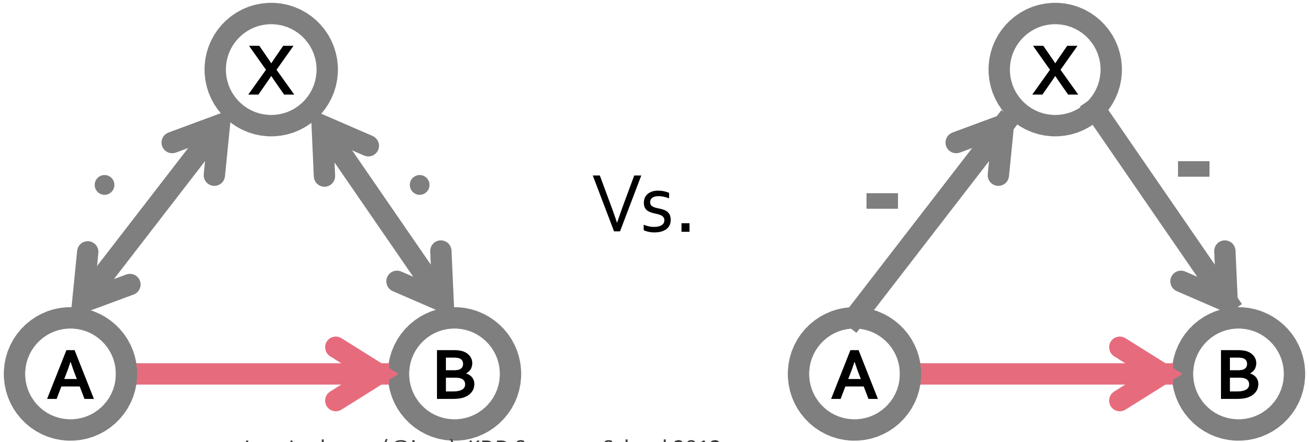
16 Types of Link Contexts

- **Link contexts:**
  - A contextualized link is a triple (A,B;X) such that directed A-B link forms after there is a two-step semi-path A-X-B
  - A-X and B-X links can have either direction and either sign: 16 possible types



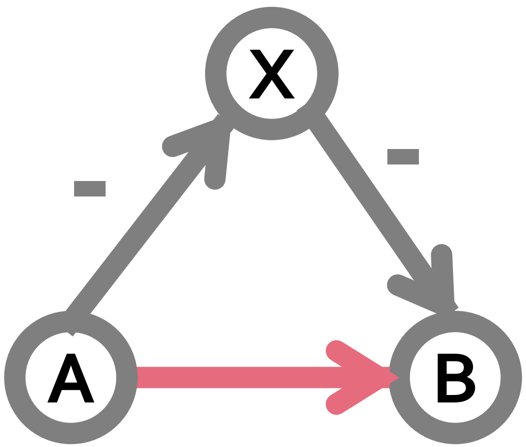
Heterogeneity in Linking Behavior

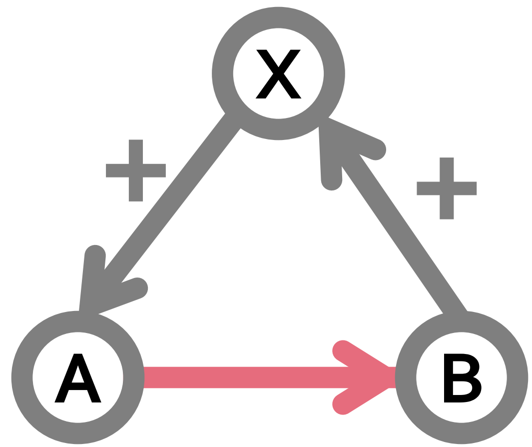
- Different users make signs differently:
  - Generative baseline (frac. of + given by A)
  - Receptive baseline (frac. of + received by B)
- **How do different link contexts cause users to deviate from baselines?**
- **Surprise**: How much behavior of A/B deviates from **baseline** when they are in **context**



Status: Two Examples

- Two basic examples:





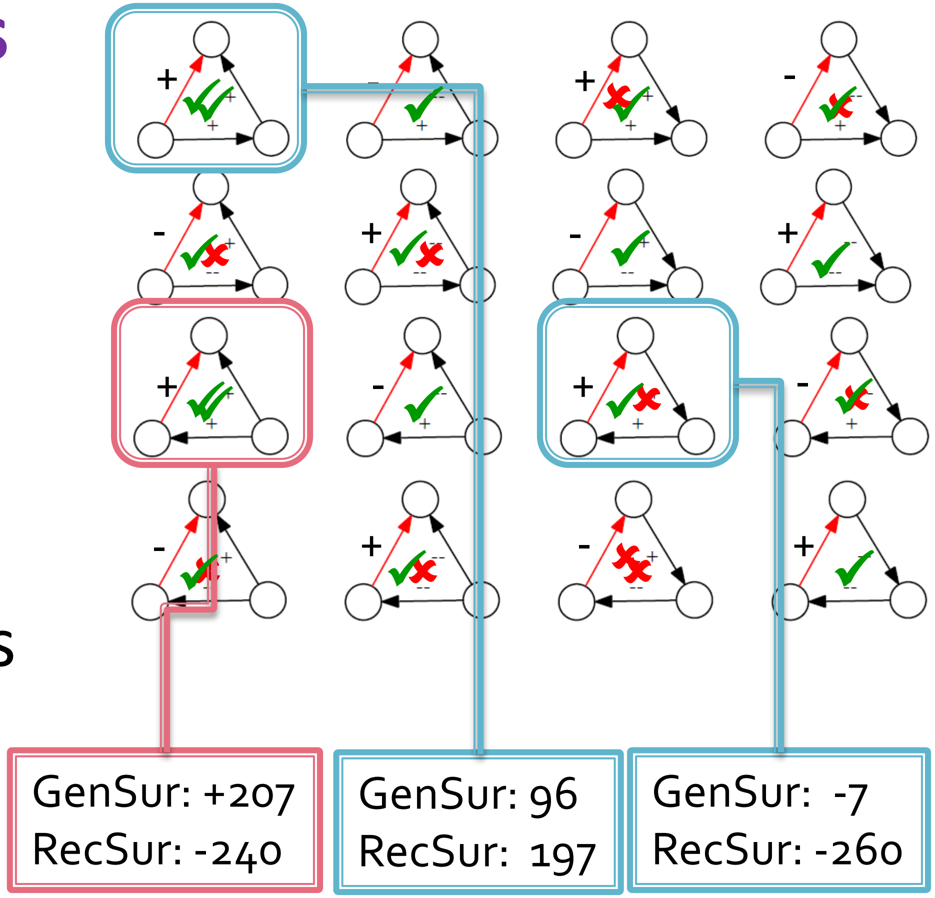
More **negative** than gen. baseline of A
More **negative** than rec. baseline of B

More **negative** than gen. baseline of A
More **negative** than rec. baseline of B

Status: Summary of results

Out of 16 triad contexts

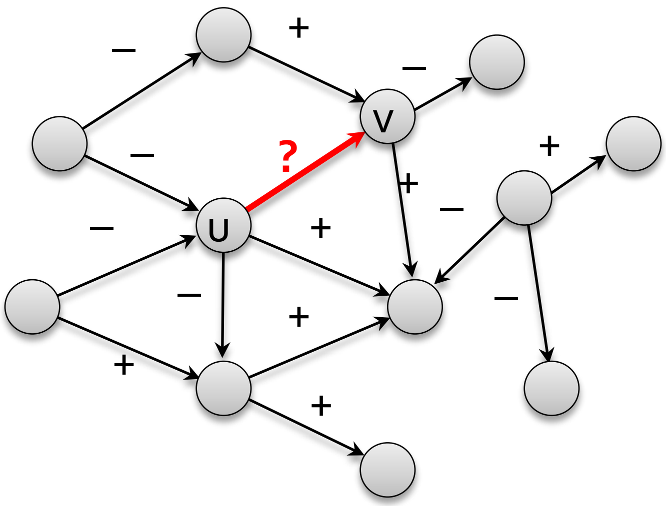
- **Generative surprise:**
  - Balance-consistent: 8
  - Status-consistent: 14
  - Both mistakes of status happen when A and B have low status
- **Receptive surprise:**
  - Status-consistent: 13
  - Balance-consistent: 7



Predicting Edge Signs

Edge sign prediction problem

- Given a network and signs on all but one edge, predict the missing sign



Machine Learning formulation:

- Predict sign of edge (u,v)
- Class label:
  - +1: positive edge
  - -1: negative edge
- Learning method:
  - Logistic regression

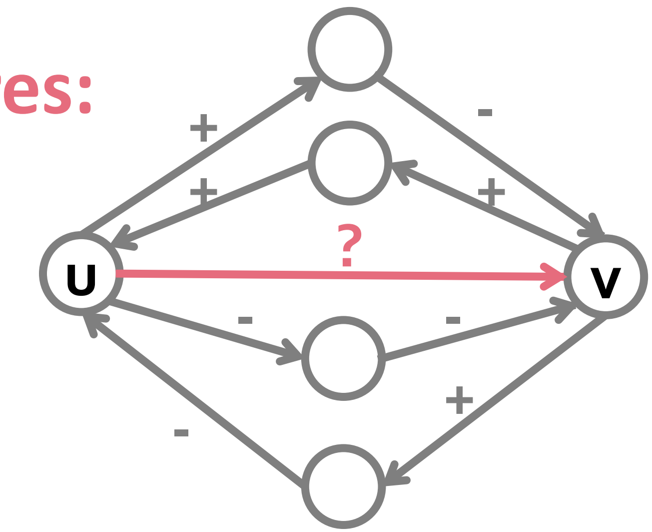
$$P(+|x) = \frac{1}{1 + e^{-(b_0 + \sum_i^n b_i x_i)}}$$

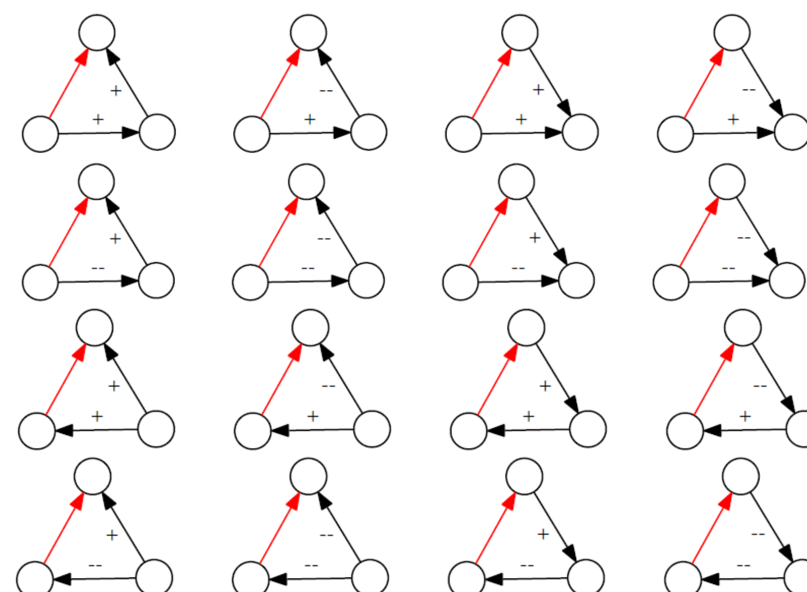
- Dataset:
  - Original: 80% +edges
  - Balanced: 50% +edges
- Evaluation:
  - Accuracy and ROC curves
- Features for learning:
  - Next slide

Features for Learning

For each edge (u,v) create features:

- Triad counts (16):
  - Counts of signed triads edge u→v takes part in
- Degree (7 features):
  - Signed degree:
    - $d^+_{out}(u)$, $d^-_{out}(u)$, $d^+_{in}(v)$, $d^-_{in}(v)$
  - Total degree:
    - $d_{out}(u)$, $d_{in}(v)$
  - Embeddedness of edge (u,v)

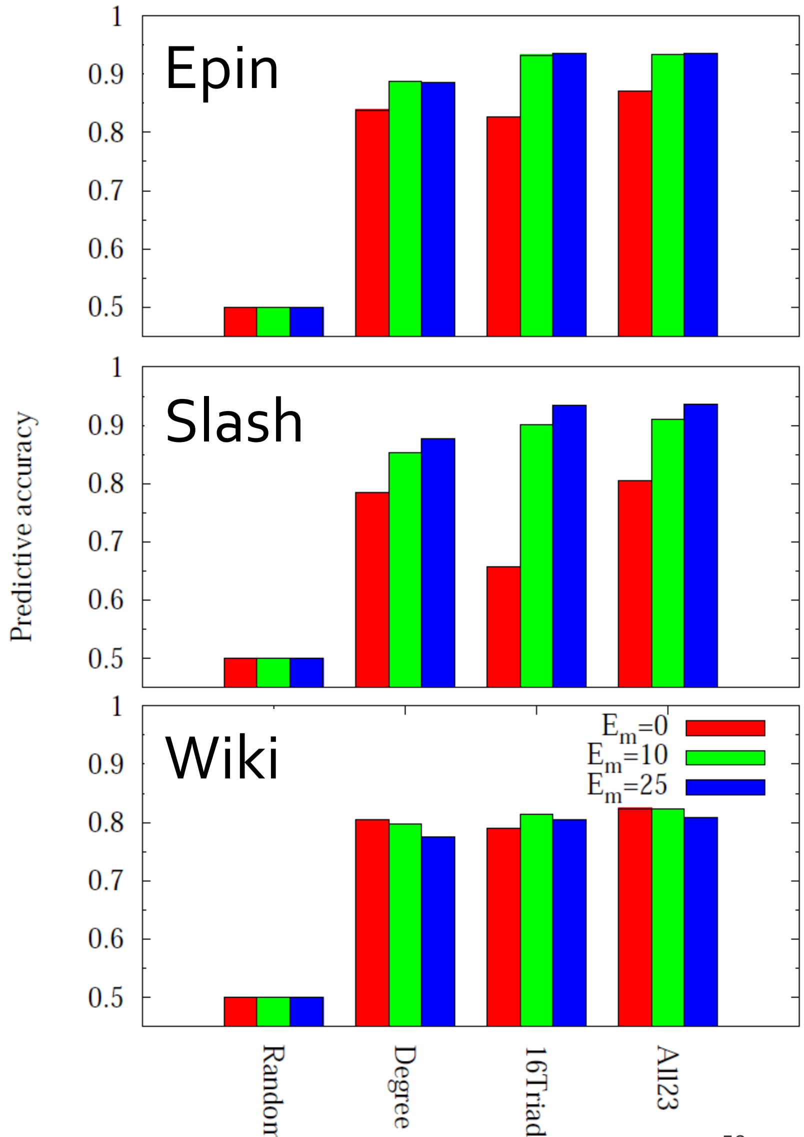




Edge Sign Prediction

[WWW '10]

- **Error rates:**
  - Epinions: 6.5%
  - Slashdot: 6.6%
  - Wikipedia: 19%
- Signs can be modeled from network structure alone
- Performance degrades for less embedded edges
- Wikipedia is harder:
  - Votes are publicly visible



Generalization

- Do people use these very different linking systems by obeying the same principles?
  - Generalization of results across the datasets?
    - Train on row "dataset", predict on "column"

All23	Epinions	Slashdot	Wikipedia
Epinions	0.9342	0.9289	0.7722
Slashdot	0.9249	0.9351	0.7717
Wikipedia	0.9272	0.9260	0.8021

- Nearly **perfect generalization** of the models even though networks come from very different applications

Final Remarks

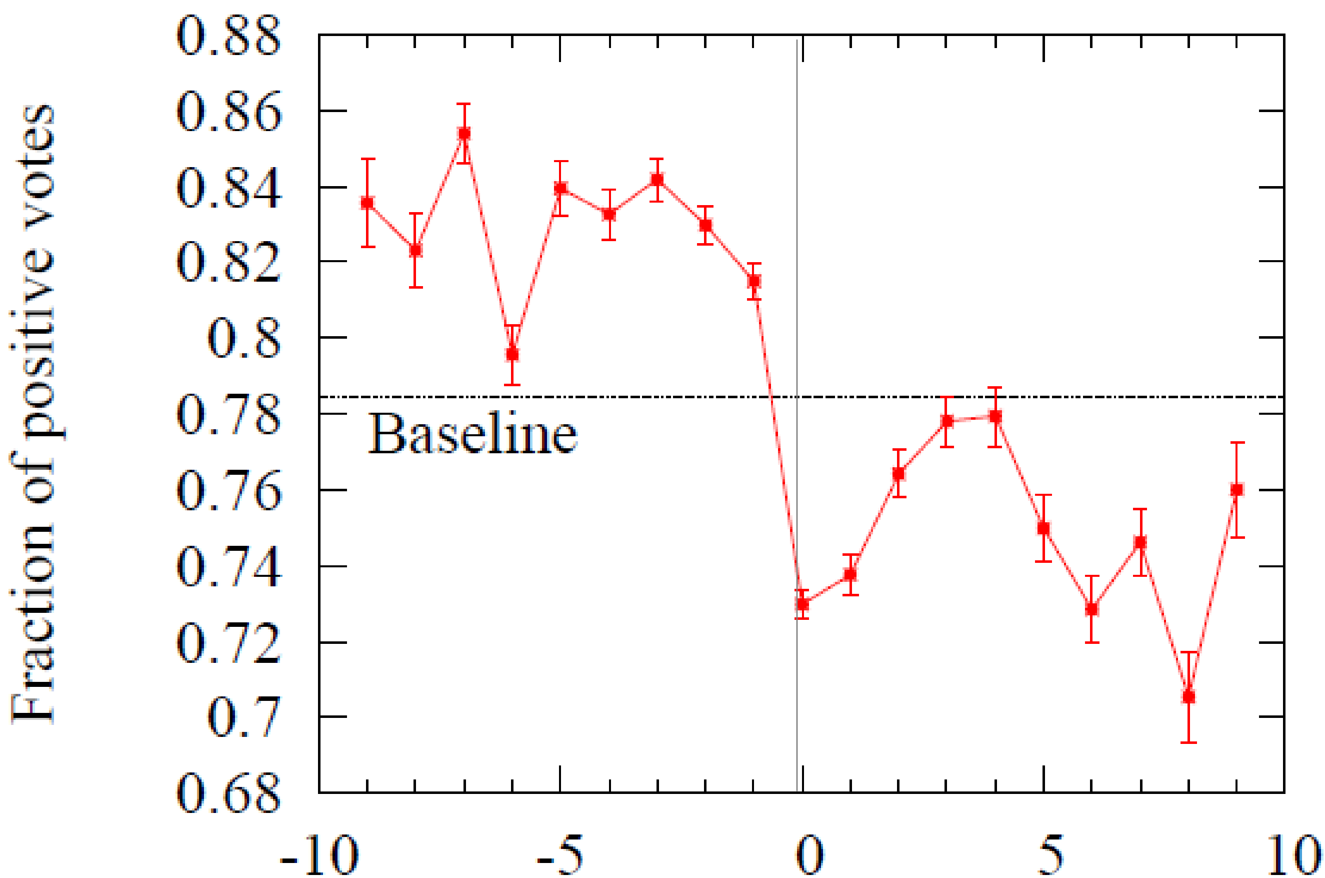
- Signed networks provide insight into how social computing systems are used:
  - Status vs. Balance
- **Sign of relationship can be reliably predicted from the local network context**
  - ~90% accuracy sign of the edge

Final Remarks

- More evidence that **networks are globally organized based on status**
- People use signed edges **consistently regardless of particular application**
  - Near perfect generalization of models across datasets
- **Many further directions:**
  - Status difference [ICWSM ‘10]

Final Remarks: Status

- Status difference on Wikipedia:

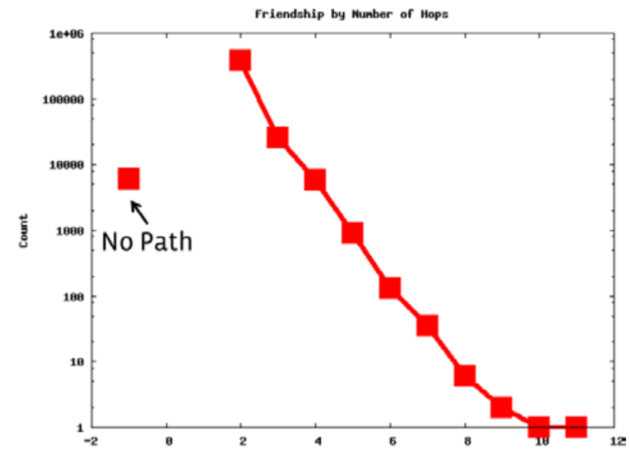


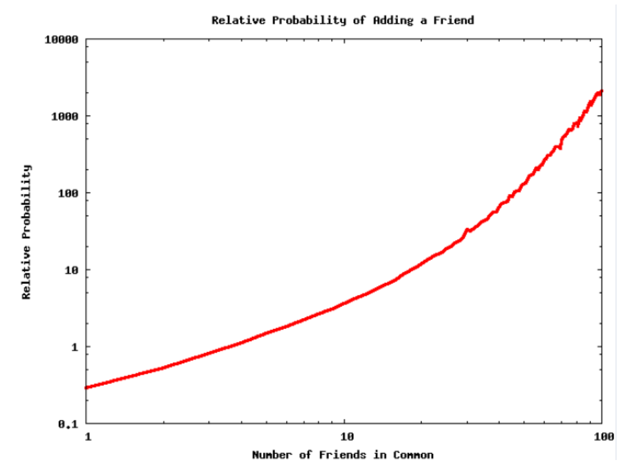
Supervised Random Walks

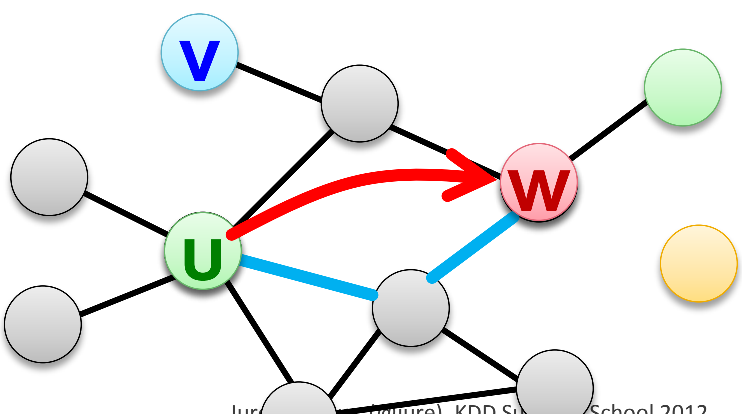
- Learning to rank nodes on a graph
- For recommending people you may know

Supervised Link Prediction

- How to learn to predict/recommend new friends in networks?
  - Facebook People You May Know
- Let's look at the data:
  - 92% of new friendships on FB are friend-of-a-friend
  - More common friends helps







Link Prediction: Challenges

- **How to learn models that combine:**
  - Network connectivity structure
  - node/edge metadata
- **Class imbalance:**
  - You only have 1,000 (out of 800M possible) friends on Facebook
  - Even if we limit prediction to friends-of-friends a typical Facebook person has 20,000 FoFs

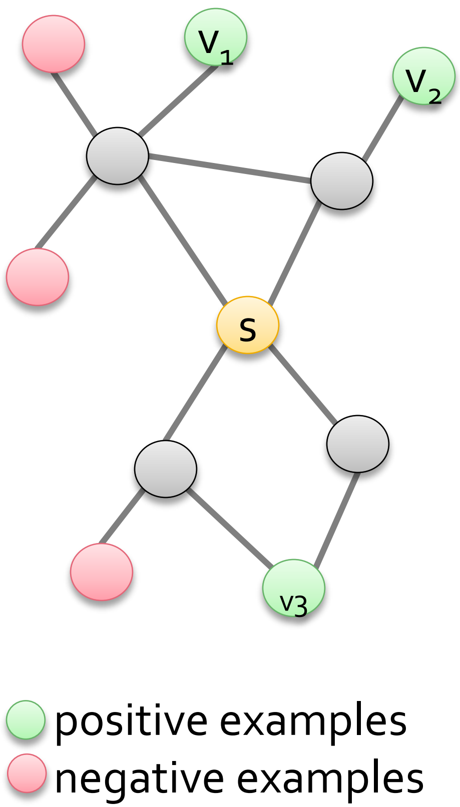
Link Prediction: Solution

- **Want to predict new Facebook friends!**
- **Combining link information and metadata:**
  - PageRank is great with network structure
  - Logistic regression is great for classification

    Lets combine the two!
- **Class imbalance:**
  - Formulate prediction task a ranking problem
- **Supervised Random Walks**
  - Supervised learning to rank nodes on a graph using PageRank

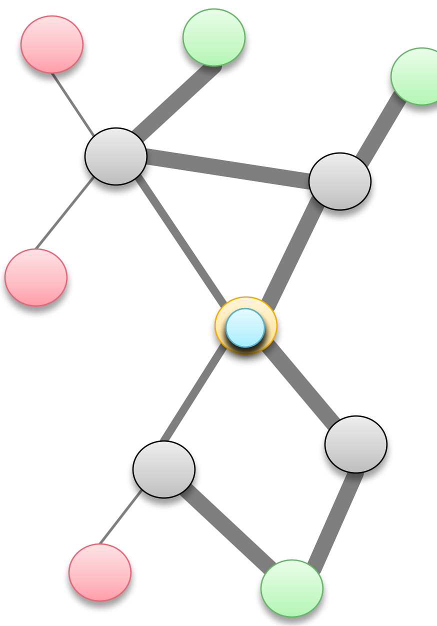
Supervised Link Prediction

- **Recommend a list of possible friends**
- Supervised machine learning setting:
  - Training example:
    - For every node s have a list of nodes she will create links to $\{v_1, \ldots, v_k\}$
      - *E.g.*, use FB network from May 2011 and $\{v_1, \ldots, v_k\}$ are the new friendships you created since then
  - Problem:
    - For a given node s **learn to rank** nodes $\{v_1, \ldots, v_k\}$ higher than other nodes in the network
- **Supervised Random Walks** based on work by Agarwal&Chakrabarti



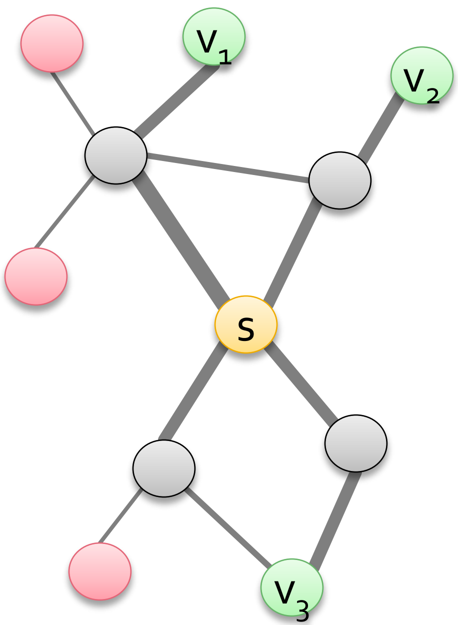
Supervised Link Prediction

- How to combine node/edge attributes and the network structure?
  - Learn a strength of each edge based on:
    - Profile of user u, profile of user v
    - Interaction history of u and v
  - Do a PageRank-like random walk from s to measure the "proximity" between s and other nodes
  - Rank nodes by their "proximity" (i.e., visiting prob.)



Supervised Random Walks

- Let s be the center node
- Let $f_w(u,v)$ be a function that assigns a **strength to each edge:**
  $$a_{uv} = f_w(u,v) = exp(-w^T \Psi_{uv})$$
  - Ψ_{uv} is a feature vector
    - Features of nodes u and v
    - Features of edge (u,v)
  - **w is the parameter vector we want to learn**
- Do Random Walk with Restarts from s where transitions are according to edge strengths
- **How to learn $f_w(u,v)$?**



Personalized PageRank

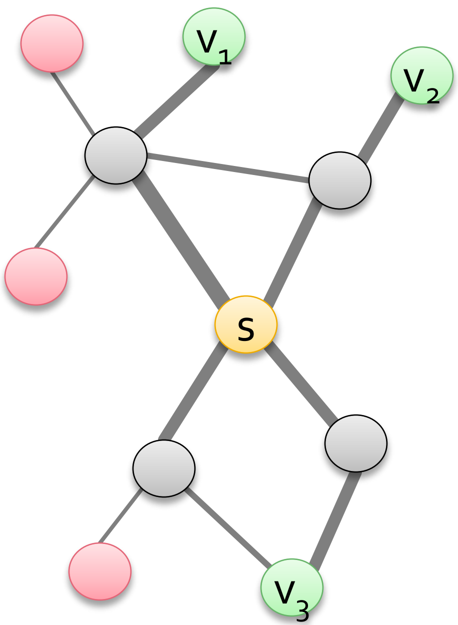
- **Random walk transition matrix:**

$$Q'_{uv} = \begin{cases} \frac{a_{uv}}{\sum_w a_{uw}} & \text{if } (u,v) \in E, \\ 0 & \text{otherwise} \end{cases}$$

- **PageRank transition matrix:**

$$Q_{ij} = (1-\alpha)Q'_{ij} + \alpha \mathbf{1}(j = s)$$

  - with prob. α jump back to s



- Compute PageRank vector: $p = p^T Q$
- Rank nodes by p_u

The Optimization Problem

- Each node u has a score p_u
- Destination nodes $D = \{v_1, \ldots, v_k\}$
- No-link nodes $L = \{\text{the rest}\}$
- **What do we want?**

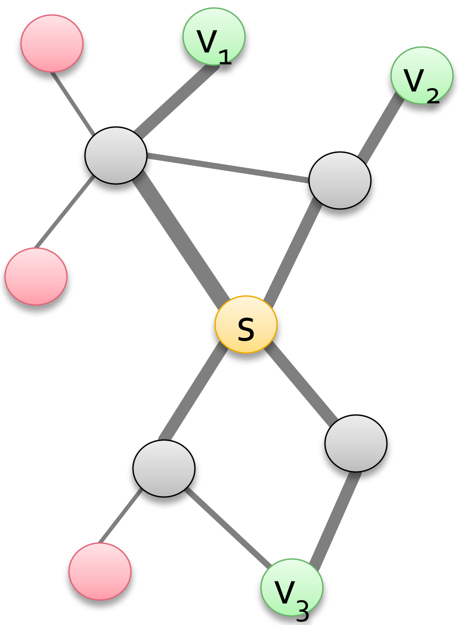
  Want to find w such that $p_l < p_d$

$$\min_{w} F(w) = ||w||^2$$

such that

$$\forall\, d \in D, l \in L:\ p_l < p_d$$

  - Hard constraints, make them soft

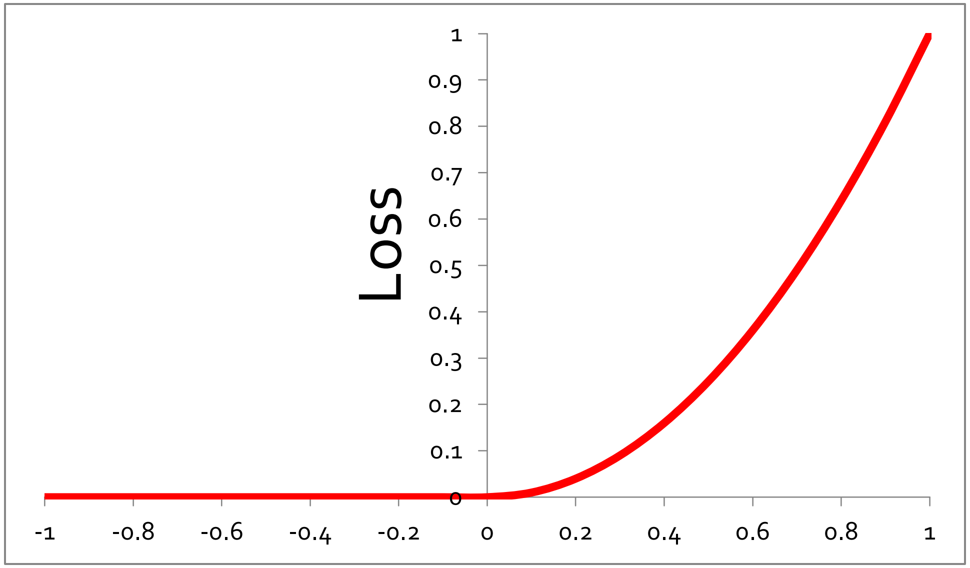


Making constraints soft

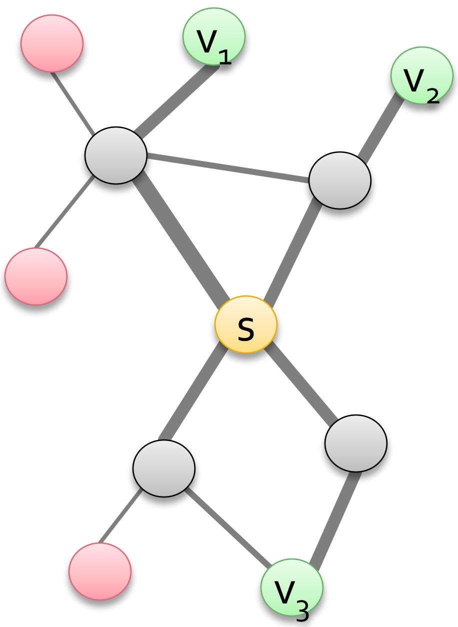
- **Want to minimize:**

$$\min_w F(w) = ||w||^2 + \lambda \sum_{ld} h(p_l - p_d)$$

  - **Loss:** $h(x) = 0$ if $x < 0$, x^2 else



$p_l < p_d$ $\quad$ $p_l = p_d$ $\quad$ $p_l > p_d$



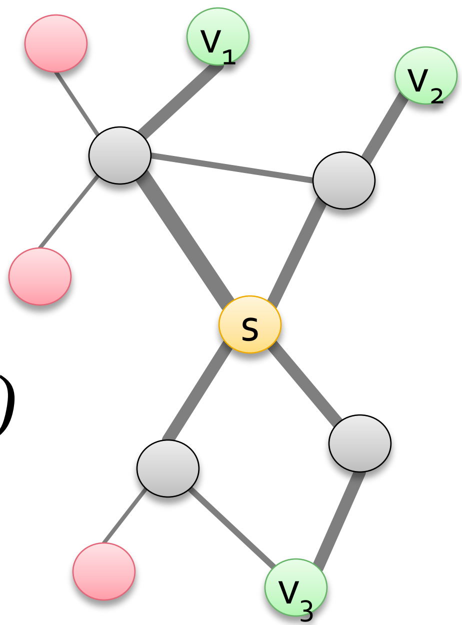
[WSDM '11]

Solving the problem: Intuition

- How to minimize *F*?

$$\min_{w} F(w) = ||w||^2 + \lambda \sum_{ld} h(p_l - p_d)$$

- p_l and p_d depend on w
  - Given w assign edge weights $a_{uv}=f_w(u,v)$
  - Using transition matrix $Q=[a_{uv}]$ compute PageRank scores p_u
  - Rank nodes by the PageRank score
- Want to find w such that $p_l < p_d$



[WSDM '11]

Gradient Descent

- How to minimize F?

$$\min_w F(w) = ||w||^2 + \lambda \sum_{l,d} h(p_l - p_d)$$

  - Take the derivative!

$$\frac{\partial F}{\partial w} = 2w + \sum_{l,d} \frac{\partial h(p_l - p_d)}{\partial w}$$

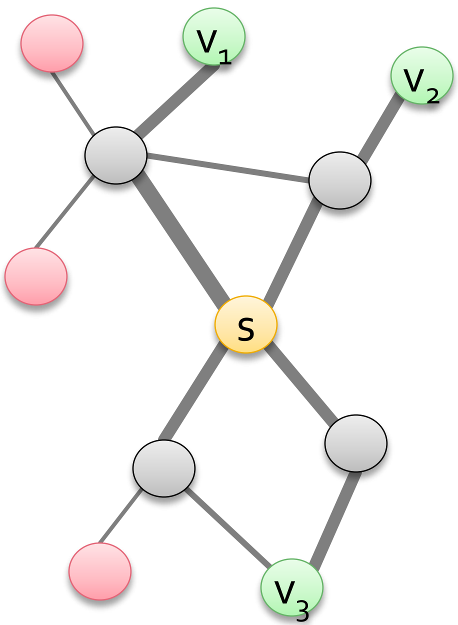
$$= 2w + \sum_{l,d} \frac{\partial h(\delta_{ld})}{\partial \delta_{ld}} \left(\frac{\partial p_l}{\partial w} - \frac{\partial p_d}{\partial w}\right)$$

- We know:

$$p = p^T Q \text{ i.e. } p_u = \sum_j p_j Q_{ju}$$

- So:

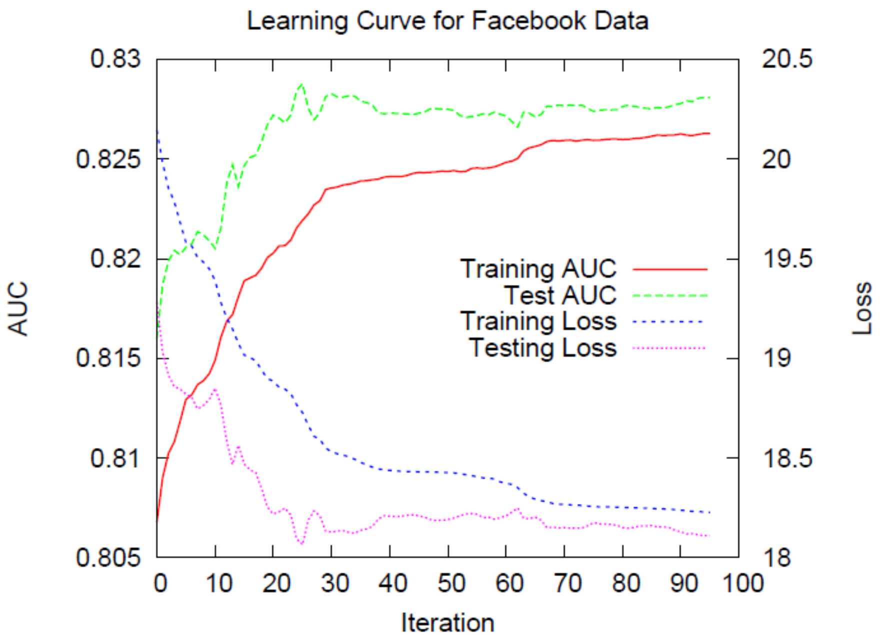
$$\frac{\partial p_u}{\partial w} = \sum_j Q_{ju} \frac{\partial p_j}{\partial w} + p_j \frac{\partial Q_{ju}}{\partial w}$$



Solve using power iteration!

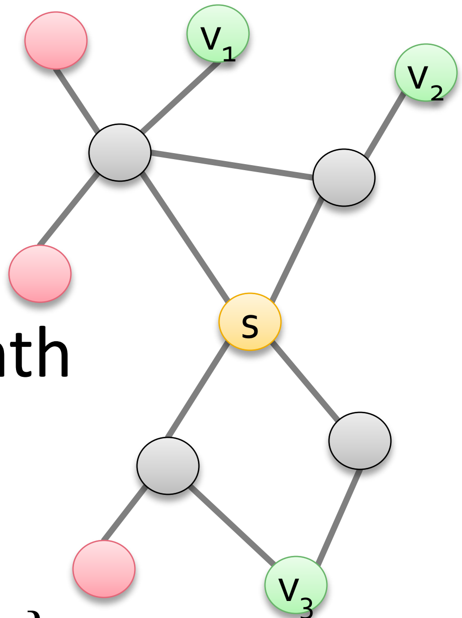
Optimizing F

- **To optimize F, use gradient based method:**
  - Pick a random starting point w_0
  - Compute the personalized PageRank vector p
  - Compute gradient with respect to weight vector w
  - Update w
    - Optimize using quasi-Newton method



Data: Facebook

- **Facebook Iceland network**
  - 174,000 nodes (55% of population)
  - Avg. degree 168
  - Avg. person added 26 new friends/month
- **For every node *s*:**
  - **Positive examples:**
    - D={ new friendships of *s* created in Nov '09 }
  - **Negative examples:**
    - L={ other nodes *s* did not create new links to }
  - Limit to friends of friends
    - on avg. there are 20k FoFs (max 2M)!



Experimental setting

- **Node and edge features:**
  - Node:
    - Age, Gender, Degree
  - Edge:
    - Edge age, Communication, Profile visits, Co-tagged photos
- **Baselines:**
  - Decision trees and logistic regression:
    - Above features + 10 network features (PageRank, common friends, …)
- **Evaluation:**
  - AUC and Precision at Top20

Results: Facebook Iceland

- **Facebook:** **predict future friends**
  - Adamic-Adar already works great
  - Logistic regression also strong
  - SRW gives slight improvement

Learning Method	AUC	Prec@20
Random Walk with Restart	0.81725	6.80
Adamic-Adar	0.81586	7.35
Common Friends	0.80054	7.35
Degree	0.58535	3.25
DT: Node features	0.59248	2.38
DT: Network features	0.76979	5.38
DT: Node+Network	0.76217	5.86
DT: Path features	0.62836	2.46
DT: All features	0.72986	5.34
LR: Node features	0.54134	1.38
LR: Network features	0.80560	7.56
LR: Node+Network	0.80280	7.56
LR: Path features	0.51418	0.74
LR: All features	0.81681	7.52
SRW: one edge type	**0.82502**	6.87
SRW: multiple edge types	**0.82799**	**7.57**

Results: Co-authorship

- **Arxiv Hep-Ph collaboration network:**
  - Poor performance of unsupervised methods
  - Logistic regression and decision trees don't work to well
  - SRW gives 10% boos in Prec@20

Learning Method	AUC	Prec@20
Random Walk with Restart	0.63831	3.41
Adamic-Adar	0.60570	3.13
Common Friends	0.59370	3.11
Degree	0.56522	3.05
DT: Node features	0.60961	3.54
DT: Network features	0.59302	3.69
DT: Node+Network	0.63711	3.95
DT: Path features	0.56213	1.72
DT: All features	0.61820	3.77
LR: Node features	0.64754	3.19
LR: Network features	0.58732	3.27
LR: Node+Network	0.64644	3.81
LR: Path features	0.67237	2.78
LR: All features	0.67426	3.82
SRW: one edge type	0.69996	4.24
SRW: multiple edge types	**0.71238**	**4.25**

References

- Kronecker Graphs: An approach to modeling networks by J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, Z. Ghahramani. *Journal of Machine Learning Research (JMLR)* 11(Feb):985-1042, 2010.
- Multiplicative Attribute Graph Model of Real-World Networks by M. Kim, J. Leskovec. *Internet Mathematics 8(1-2) 113--160* , 2012.
- Modeling Social Networks with Node Attributes using the Multiplicative Attribute Graph Model by M. Kim, J. Leskovec. *Conference on Uncertainty in Artificial Intelligence (UAI)*, 2011.
- Latent Multi-group Membership Graph Model by M. Kim, J. Leskovec. *International Conference on Machine Learning (ICML)*, 2012.
- Supervised Random Walks: Predicting and Recommending Links in Social Networks by L. Backstrom, J. Leskovec. *ACM International Conference on Web Search and Data Mining (WSDM)*, 2011.

References

- Signed Networks in Social Media by J. Leskovec, D. Huttenlocher, J. Kleinberg. *ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 2010.
- Predicting Positive and Negative Links in Online Social Networks by J. Leskovec, D. Huttenlocher, J. Kleinberg. *ACM WWW International conference on World Wide Web (WWW)*, 2010.
- Governance in Social Media: A case study of the Wikipedia promotion process by J. Leskovec, D. Huttenlocher, J. Kleinberg. *AAAI International Conference on Weblogs and Social Media (ICWSM)*, 2010.
- Effects of User Similarity in Social Media by A. Anderson, D. Huttenlocher, J. Kleinberg, J. Leskovec.*ACM International Conference on Web Search and Data Mining (WSDM)*, 2012.

THANKS!
http://snap.stanford.edu