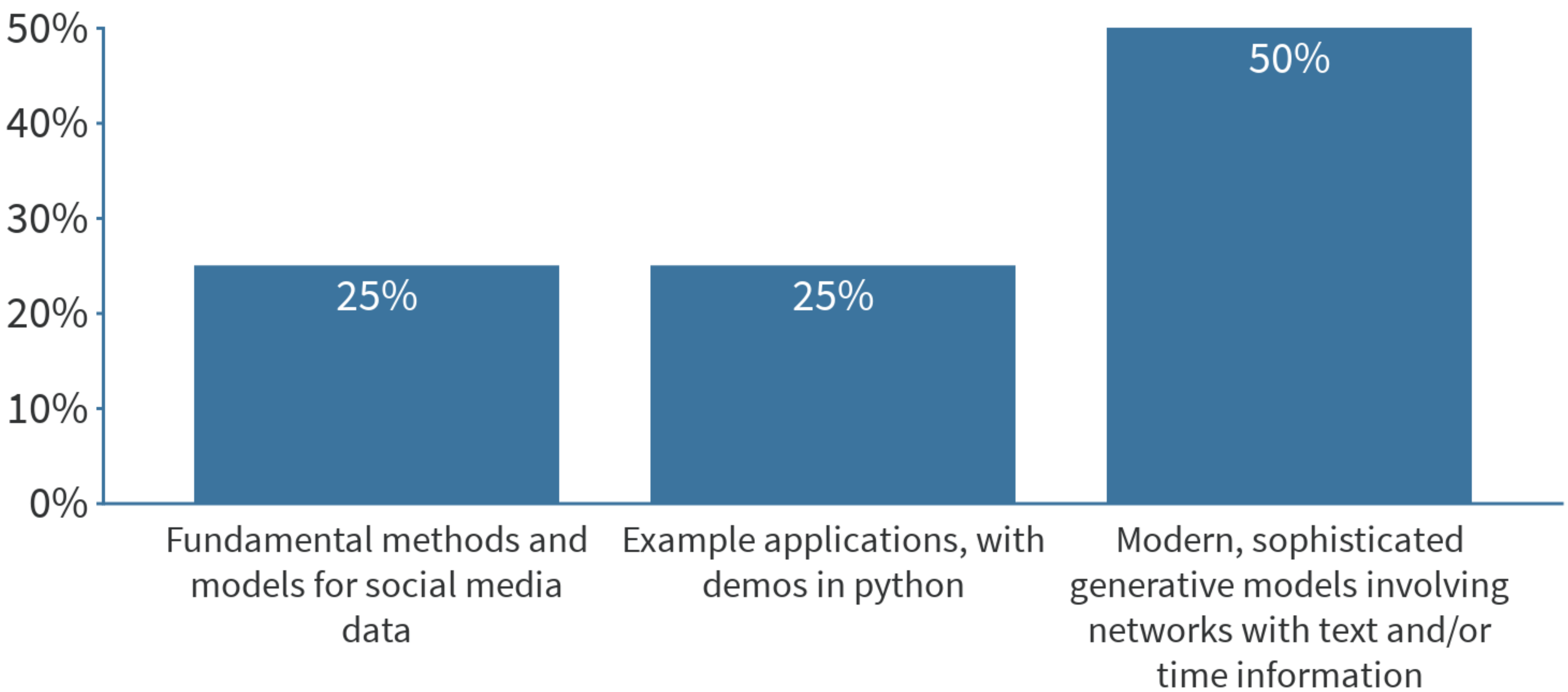
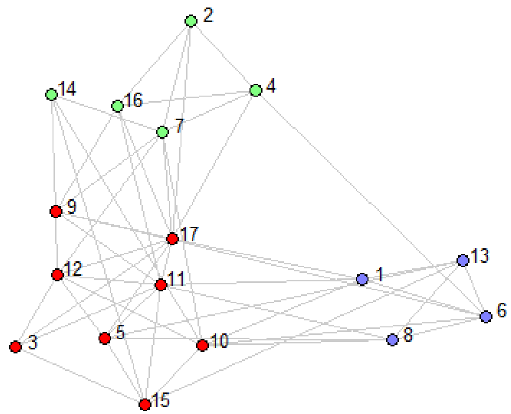
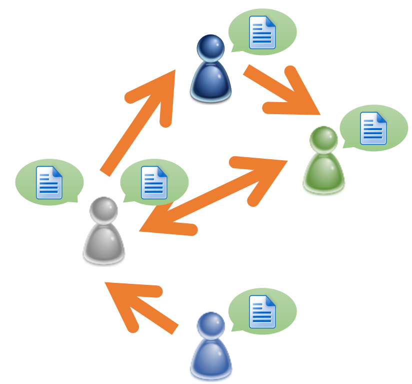
Before the tutorial starts, if you can, please vote on this poll via your smartphone, tablet, or laptop. The part of this tutorial I am most interested learning about is:

When poll is active, respond at **PollEv.com/jamesfoulds656** Text **JAMESFOULDS656** to **37607** once to join







Generative Models for Social Media Analytics: Networks, Text, and Time

Kevin S. Xu (University of Toledo)

James R. Foulds (University of Maryland-Baltimore County)

ICWSM 2018 Tutorial

About Us

Kevin S. Xu

- Assistant professor at University of Toledo
- 3 years research experience in industry
- Research interests:
  - Machine learning
  - Statistical signal processing
  - Network science
  - Wearable data analytics

James R. Foulds

- Assistant professor at University of Maryland-Baltimore County
- Research interests:
  - Bayesian modeling
  - Social networks
  - Text
  - Latent variable models

Social media data

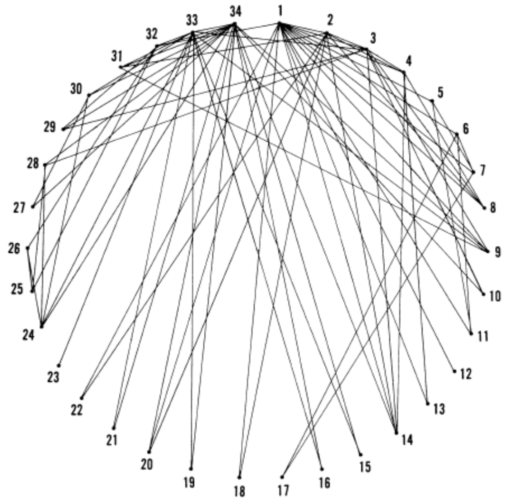
- Content
  - Text
  - Images
  - Video
- Relations
  - Friendships/follows
  - Likes/reactions
  - Tags
  - Re-tweets
- User attributes
  - Location
  - Age
  - Interests

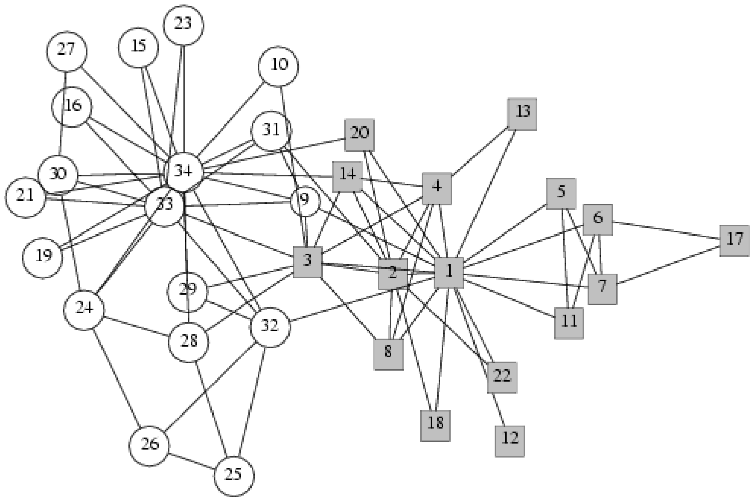
Outline

- Mathematical representations and generative models for social networks
    - Introduction to generative approach
    - Connections to sociological principles
- Fitting generative social network models to data
    - Application scenarios with demos
    - Model selection and evaluation
- Rich generative models for social media data
    - Network models augmented with text and dynamics
    - Case studies on social media data

Social networks as graphs

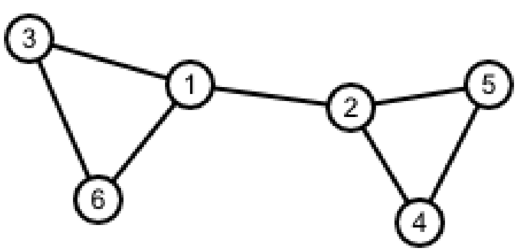
- A social network can be represented by a graph $G = (V, E)$
  - V: vertices, nodes, or actors typically representing people
  - E: edges, links, or ties denoting relationships between nodes
  - Directed graphs used to represent asymmetric relationships
- Graphs have no natural representation in a geometric space
  - Two identical graphs drawn differently
  - Moral: visualization provides very limited analysis ability
  - How do we model and analyze social network data?



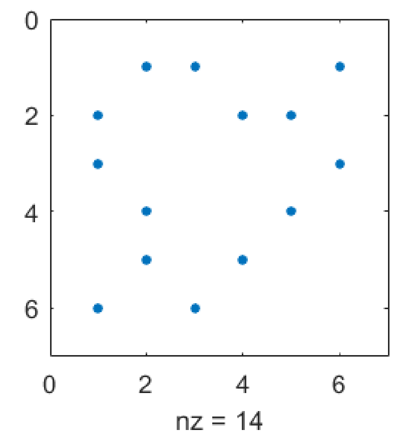


Matrix representation of social networks

- Represent graph by $n \times n$ adjacency matrix or sociomatrix $\mathbf{Y}$
  - $y_{ij} = 1$ if there is an edge between nodes i and j
  - $y_{ij} = 0$ otherwise



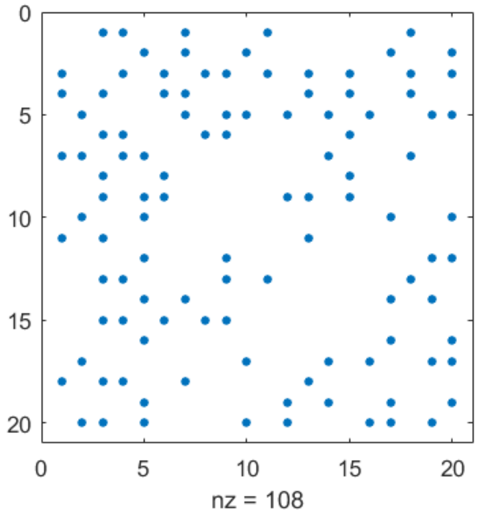
$$\mathbf{Y} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

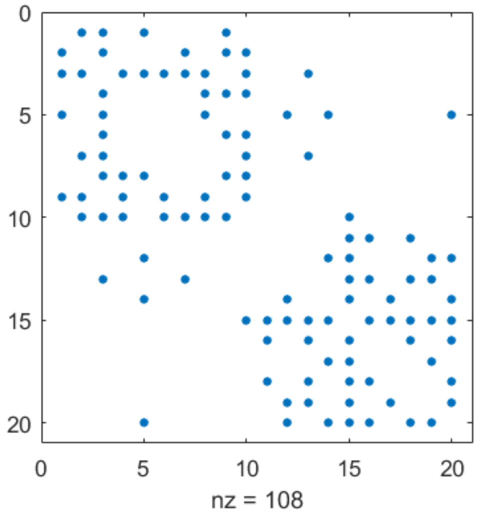


- Easily extended to directed and weighted graphs

Adjacency matrix permutation invariance

- Row and column permutations to adjacency matrix do not change graph
  - Changes only ordering of nodes
  - Provided same permutation is applied to both rows and columns
- Same graph with 2 different orderings of nodes





Sociological principles related to edge formation

- Homophily or assortative mixing
  - Tendency for individuals to bond with similar others
  - Assortative mixing by age, gender, social class, organizational role, node degree, etc.
  - Results in transitivity (triangles) in social networks
    - "My friend of my friend is my friend"
- Equivalence of nodes
  - Two nodes are structurally equivalent if their relations to all other nodes are identical
    - Approximate equivalence recorded by similarity measure
  - Two nodes are regularly equivalent if their neighbors are similar (not necessarily common neighbors)

Brief history of social network models

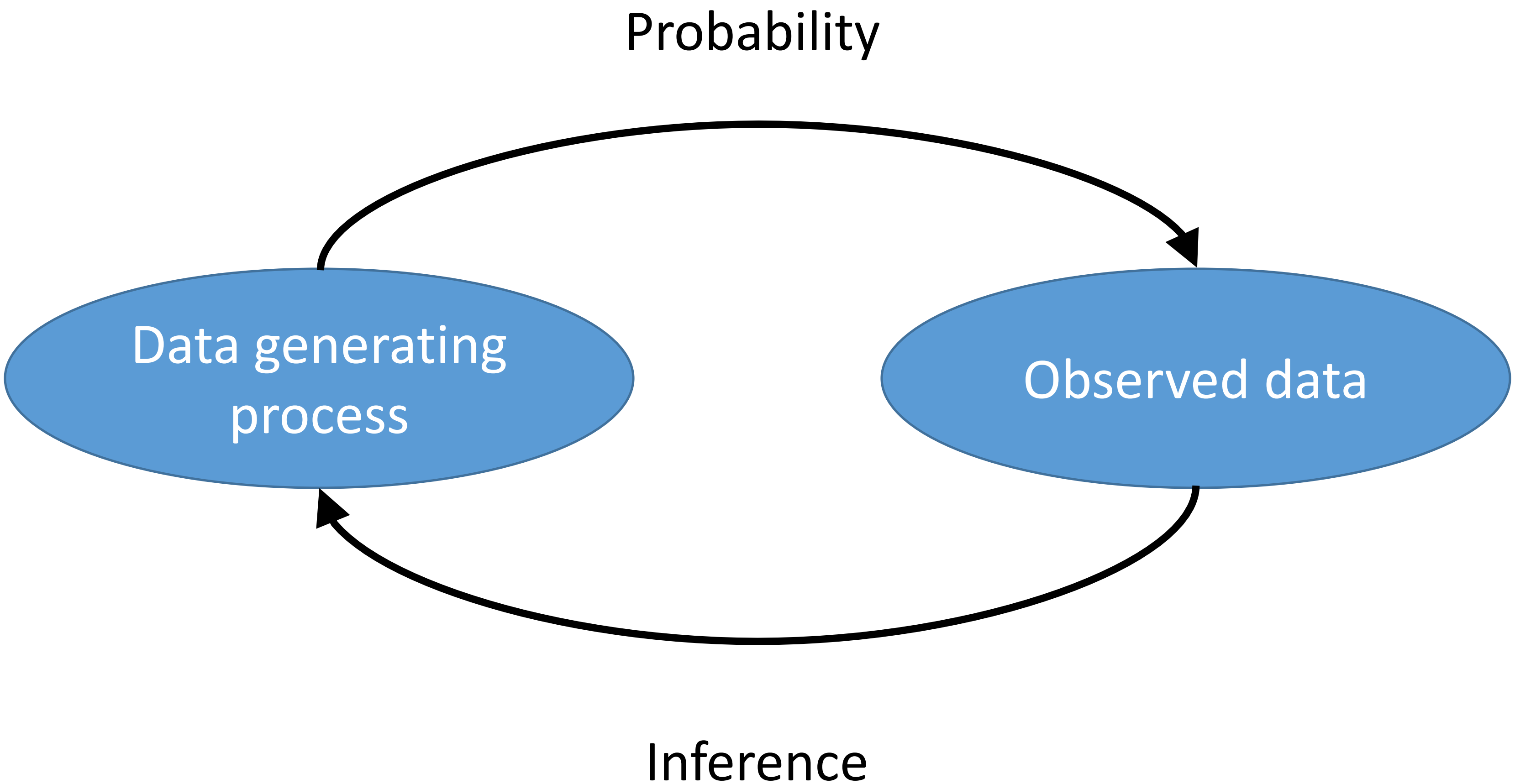
- 1930s – Graphical depictions of social networks: sociograms (Moreno)
- 1950s – Mathematical (probabilistic) models of social networks (Erdős-Rényi-Gilbert)
- 1960s – Small world / 6-degrees of separation experiment (Milgram)
- 1980s – Introduction of statistical models: stochastic block models and precursors to exponential random graph models (Holland et al., Frank and Strauss)
- 1990s – Statistical physicists weigh in: small-world models (Watts-Strogatz) and preferential attachment (Barabási-Albert)
- 2000s-today – Machine learning approaches, latent variable models

Generative models for social networks

- A generative model is one that can simulate new networks
- Two distinct schools of thought:
  - Probability models (non-statistical)
    - Typically simple, 1-2 parameters, not typically learned from data
    - Can be studied analytically
  - Statistical models
    - More parameters, latent variables
    - Learned from data via statistical estimation techniques

Probability and Inference

Mathematics/physics: Erdős-Rényi, preferential attachment,…



Statistics/machine learning: ERGMs, latent variable models…

Figure based on one by Larry Wasserman, "All of Statistics"

Probability models for networks

- Erdős-Rényi-Gilbert $G(N, p)$ model (1 parameter)
  - An edge is formed between any two nodes with equal probability p
  - 2 drawbacks with $G(N, p)$ model:
    - Does not generative networks with transitivity
    - Each node ends up with roughly same degree (number of edges)
- Watts-Strogatz small-world model (2 parameters)
  - Mechanistic construction by re-wiring edges
  - Addresses drawback #1 by creating networks with triangles and short average path lengths

Probability models for networks

- Erdős-Rényi-Gilbert $G(N, p)$ model (1 parameter)
  - An edge is formed between any two nodes with equal probability p
  - 2 drawbacks with $G(N, p)$ model:
    - Does not generative networks with transitivity
    - Each node ends up with roughly same degree (number of edges)
- Barabási-Albert model (2 parameters)
  - Mechanistic construction that grows a network from an initial "seed" using preferential attachment
  - Addresses drawback #2 by creating networks with power-law degree distributions $P(k) \propto k^{-\lambda}$

Probability models for networks

- Erdős-Rényi-Gilbert $G(N, p)$ model (1 parameter)
- Watts-Strogatz small-world model (2 parameters)
- Barabási-Albert model (2 parameters)
- Advantage: simplicity enables rigorous theoretical analysis of model properties
- Disadvantage: limited flexibility results in poor fits to data
  - Even though they are "generative", they don't generate networks that share many properties with the specific network they were fit to

Statistical models for networks

- Statistical models try to represent networks using a larger number of parameters to capture properties of a specific network
- Exponential random graph models
- Latent variable models
    - Latent space models
    - Stochastic block models
    - Mixed-membership stochastic block models
    - Latent feature models

Exponential family random graphs (ERGMs)

$$Pr(Y = y|\theta) = \frac{1}{Z(\theta)} \exp\left(\theta^\intercal S(y, \mathbf{X})\right)$$

Arbitrary sufficient statistics

Covariates (gender, age, …)

E.g. “how many males are friends with females”

Exponential family random graphs (ERGMs)

- Pros:
  - Powerful, flexible representation
  - Can encode complex theories, and do substantive social science
  - Handles covariates
  - Mature software tools available,
    e.g. ergm package for statnet

Exponential family random graphs (ERGMs)

- Cons:
  - Computationally intensive to fit to data
  - **Model degeneracy** can easily happen
    - *"a seemingly reasonable model can actually be such a bad mis-specification for an observed dataset as to render the observed data virtually impossible"*
      - **Goodreau (2007)**
- **Moral of the story:** ERGMs are powerful, but require care and expertise to perform well

Latent variable models for social networks

- Model where observed variables are dependent on a set of unobserved or latent variables
  - Observed variables assumed to be conditionally independent given latent variables
- Why latent variable models?
  - Adjacency matrix $\mathbf{Y}$ is invariant to row and column permutations
  - Aldous-Hoover theorem implies existence of a latent variable model of form

$$y_{ij} = h(\theta, z_i, z_j, \epsilon_{ij})$$

    for iid latent variables z_i and some function h

Latent variable models for social networks

- Latent variable models allow for heterogeneity of nodes in social networks
  - Each node (actor) has a latent variable $\mathbf{z}_i$
  - Probability of forming edge between two nodes is independent of all other node pairs given values of latent variables

$$p(\mathbf{Y}|\mathbf{Z},\theta) = \prod_{i \neq j} p\left(y_{ij}\middle|\mathbf{z}_i, \mathbf{z}_j, \theta\right)$$

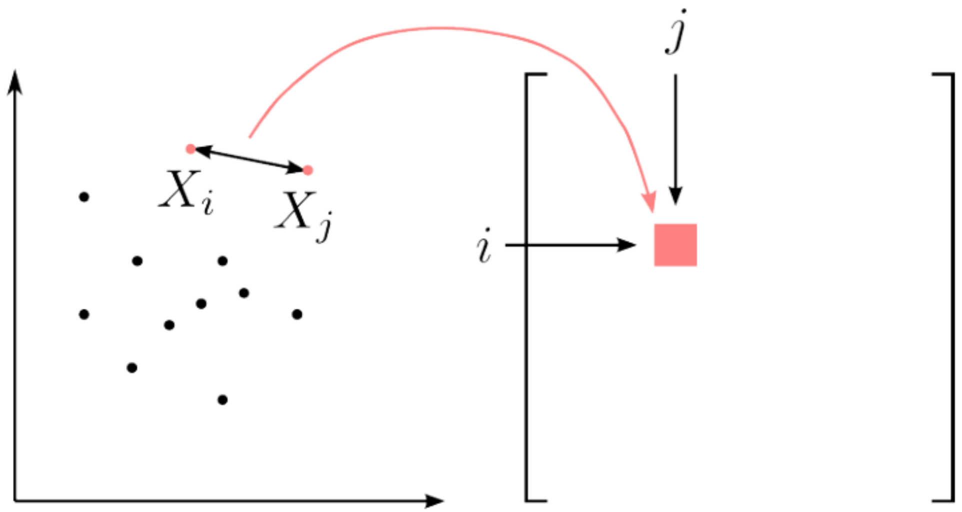
  - Ideally latent variables should provide an interpretable representation

(Continuous) latent space model

- Motivation: homophily or assortative mixing
  - Probability of edge between two nodes increases as characteristics of the nodes become more similar
- Represent nodes in an unobserved (latent) space of characteristics or “social space”
- Small distance between 2 nodes in latent space ➔ high probability of edge between nodes
  - Induces transitivity: observation of edges (i, j) and (j, k) suggests that i and k are not too far apart in latent space ➔ more likely to also have an edge

(Continuous) latent space model

- (Continuous) latent space model (LSM) proposed by Hoff et al. (2002)
  - Each node has a latent position $\mathbf{z}_i \in \mathbb{R}^d$
  - Probabilities of forming edges depend on distances between latent positions
  - Define pairwise affinities $\psi_{ij} = \theta - \left\|\mathbf{z}_i - \mathbf{z}_j\right\|_2$



Latent space model: generative process

1. Sample node positions in latent space

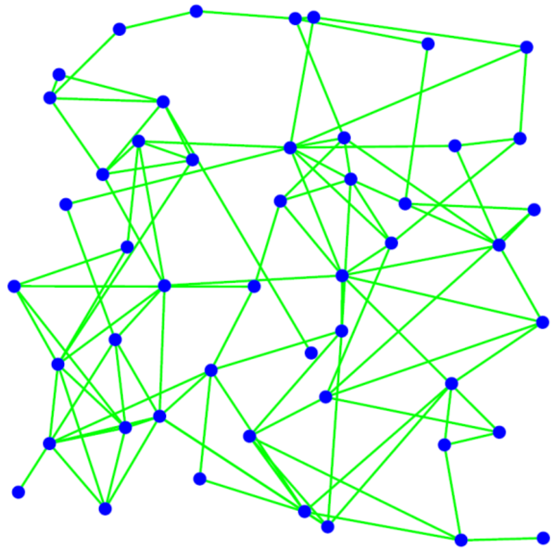
$$\mathbf{z}_i \sim \text{Gaussian}(\mathbf{0}, \kappa\mathbf{I})$$

2. Compute affinities between all pairs of nodes

$$\psi_{ij} = \theta - \|\mathbf{z}_i - \mathbf{z}_j\|_2$$

3. Sample edges between all pairs of nodes

$$P(Y_{ij} = 1|\psi_{ij}) = \sigma(\psi_{ij})$$



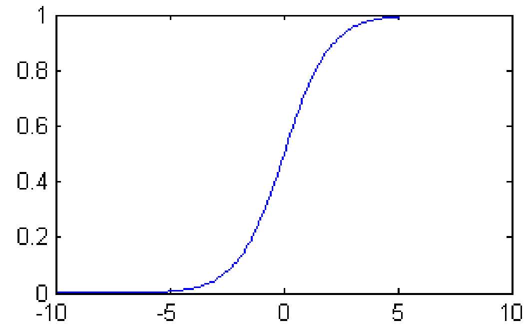


Figure due to P. D. Hoff, Modeling homophily and stochastic equivalence in symmetric relational data, NIPS 2008

Advantages and disadvantages of latent space model

- Advantages of latent space model
  - Visual and interpretable spatial representation of network
  - Models homophily (assortative mixing) well via transitivity
- Disadvantages of latent space model
  - 2-D latent space representation often may not offer enough degrees of freedom
  - Cannot model disassortative mixing (people preferring to associate with people with different characteristics)

Stochastic block model (SBM)

- First formalized by Holland et al. (1983)
- Also known as multi-class Erdős-Rényi model
- Each node has categorical latent variable $z_i \in \{1, \ldots, K\}$ denoting its class or group
- Probabilities of forming edges depend on class memberships of nodes ($K \times K$ matrix W)
  - Groups often interpreted as functional roles in social networks

$$Y \sim \begin{bmatrix} W_{11} & \cdots & W_{1K} \\ \vdots & \ddots & \vdots \\ W_{K1} & \cdots & W_{KK} \end{bmatrix}$$

Stochastic equivalence and block models

- Stochastic equivalence: generalization of structural equivalence
- Group members have identical probabilities of forming edges to members other groups
  - Can model both assortative and disassortative mixing

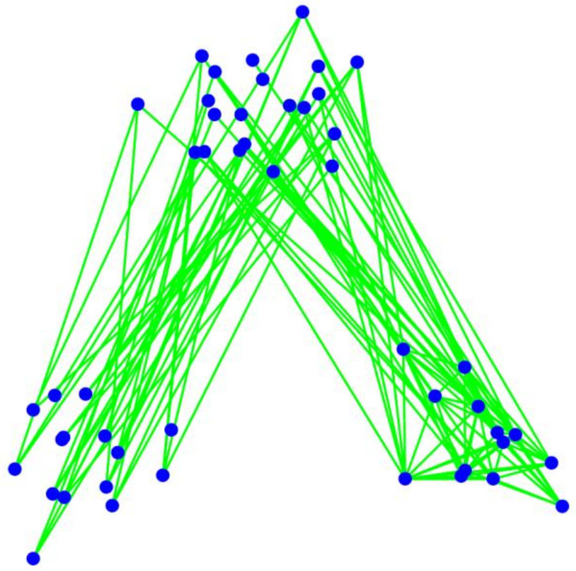
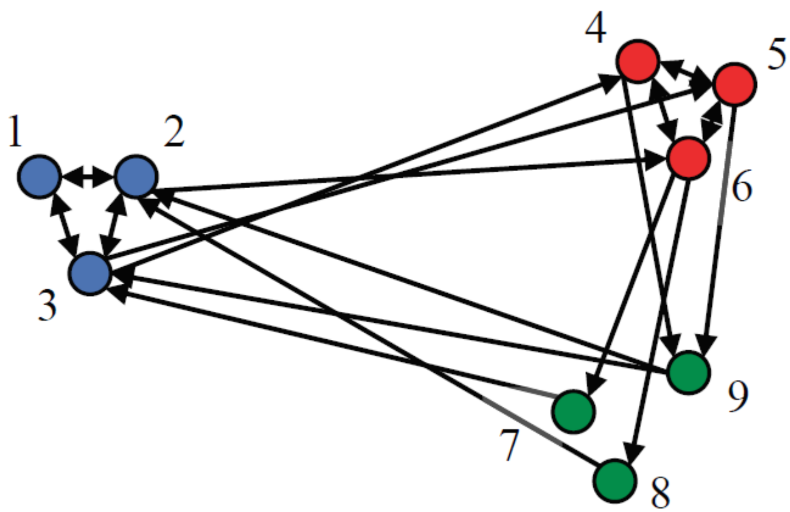


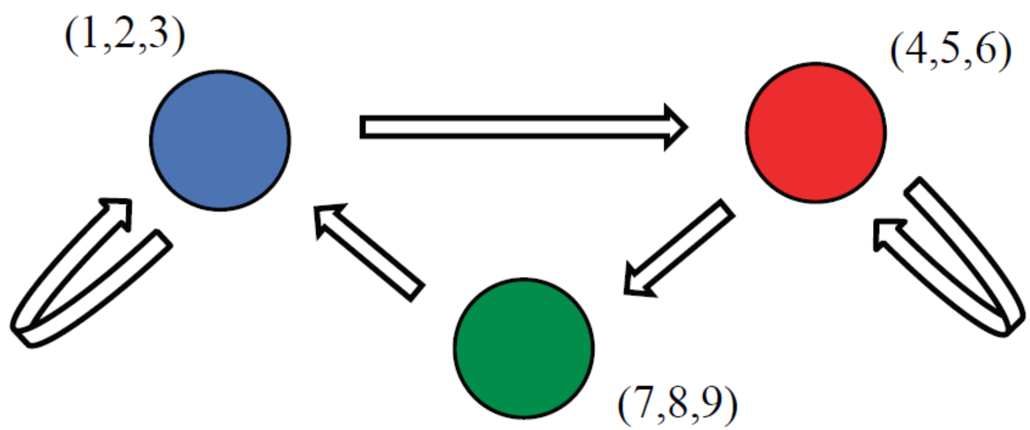
Figure due to P. D. Hoff, Modeling homophily and stochastic equivalence in symmetric relational data, NIPS 2008

Stochastic equivalence vs community detection

Original graph



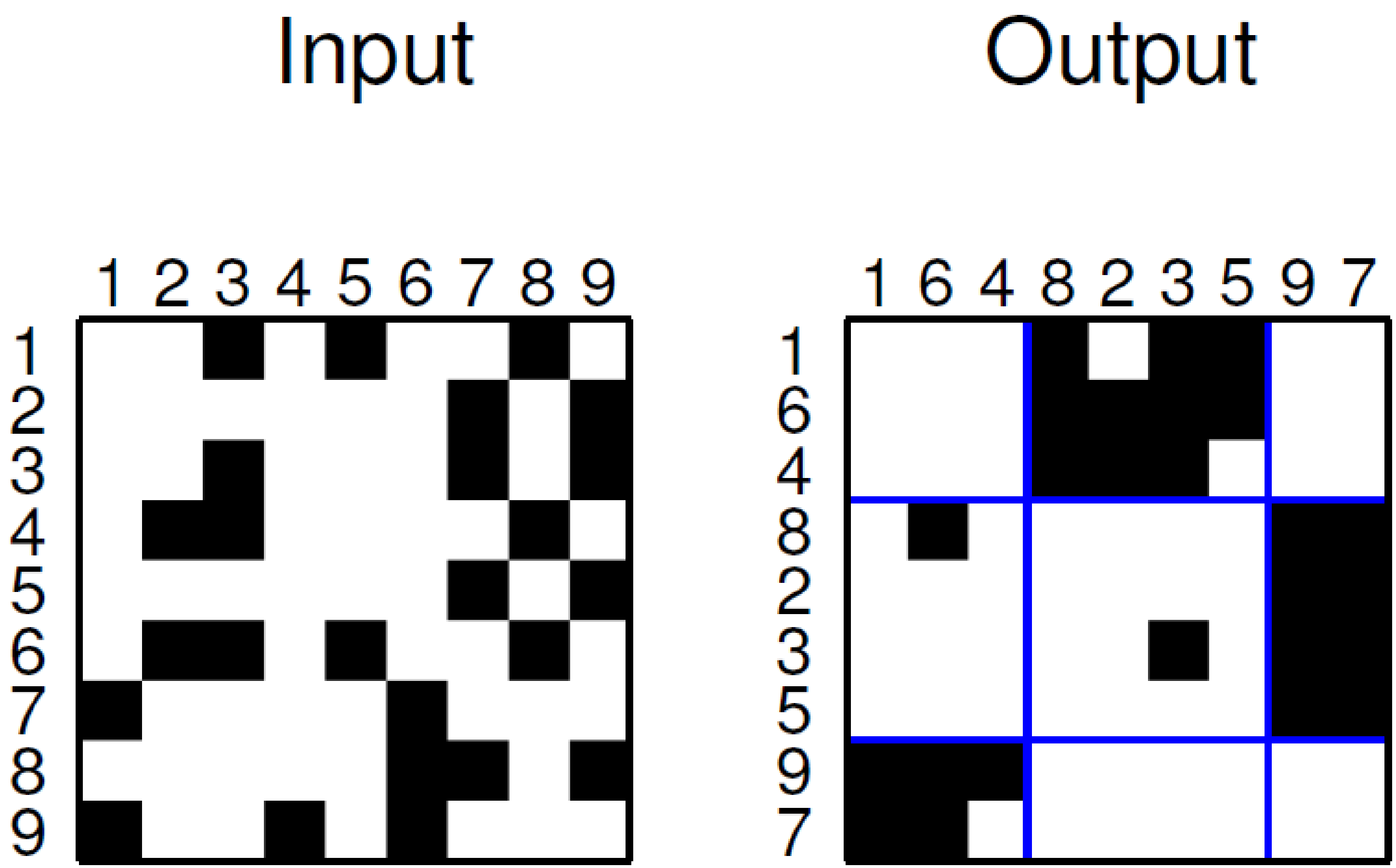
Blockmodel



Stochastically equivalent, but
are not densely connected

Figure due to Goldenberg et al. (2009) - Survey of Statistical Network Models, Foundations and Trends

Reordering the matrix to show the inferred block structure



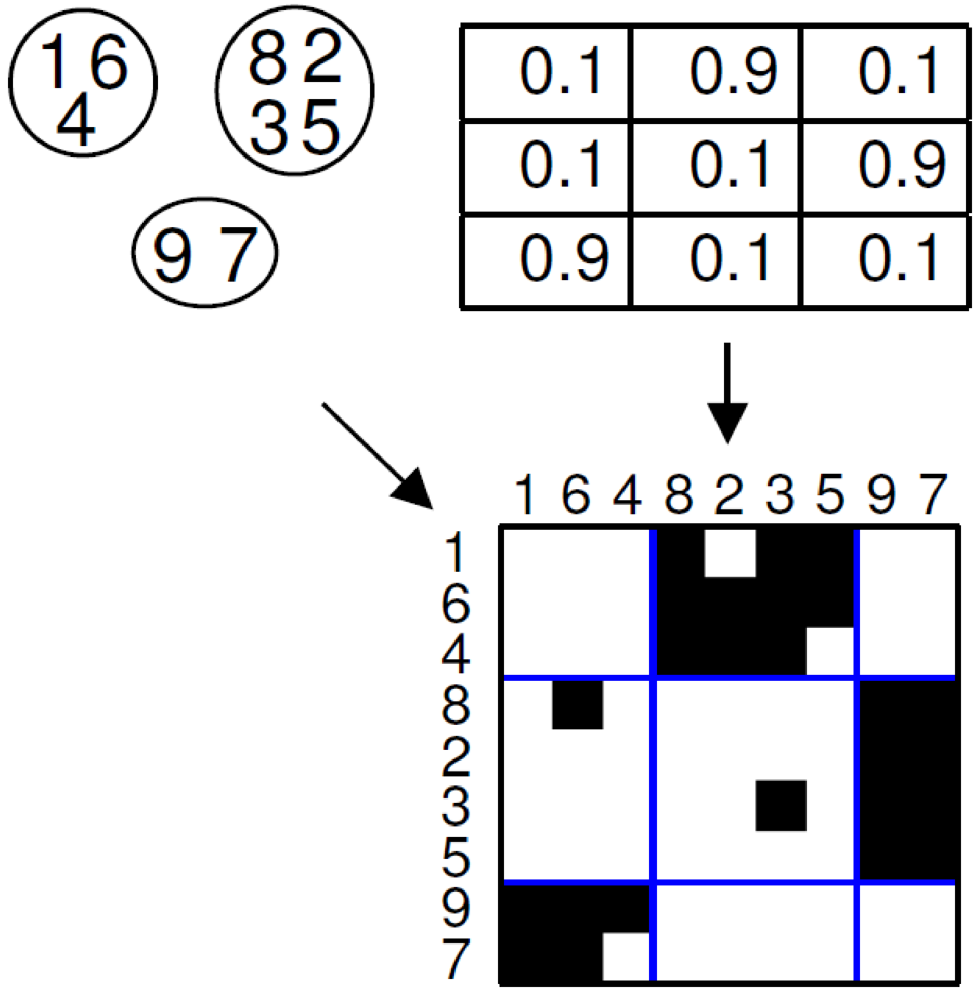
Kemp, Charles, et al. "Learning systems of concepts with an infinite relational model." AAAI. Vol. 3. 2006.

Model structure

Latent groups Z

Interaction matrix W

(probability of an edge from block k to block k')



Kemp, Charles, et al. "Learning systems of concepts with an infinite relational model." AAAI. Vol. 3. 2006.

Stochastic block model generative process

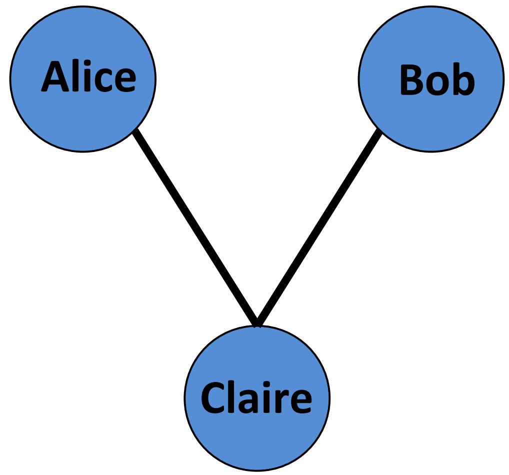
$W_{kk'}$: Probability that a node in group k connects to a node in k'

z_i: Latent group assignment for node i

For each pair of nodes (i, j)

$$Y_{ij} \sim \text{Bernoulli}(W_{z_i, z_j})$$

Stochastic block model
Latent representation



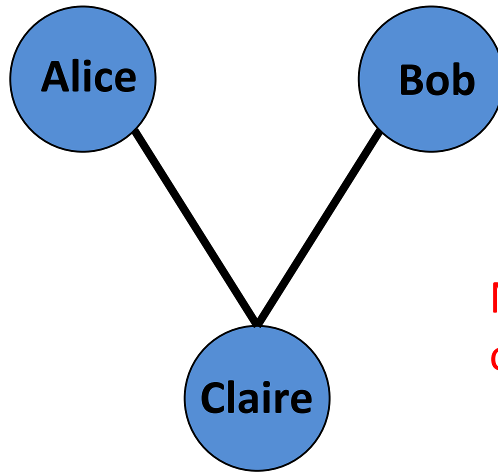
Nodes assigned to only one latent group.

Not always an appropriate assumption

$\mathbf{Z} =$

	Running	Dancing	Fishing
Alice	1		
Bob			1
Claire		1	

Mixed membership stochastic blockmodel (MMSB)



Nodes represented by **distributions** over latent groups (roles)

$\mathbf{Z} =$

	Running	Dancing	Fishing
Alice	0.4	0.4	0.2
Bob	0.5	0.5	
Claire	0.1		0.9

Airoldi et al., (2008)

Mixed membership stochastic blockmodel (MMSB)

$\pi^{(i)}$: Mixed membership vector for node i

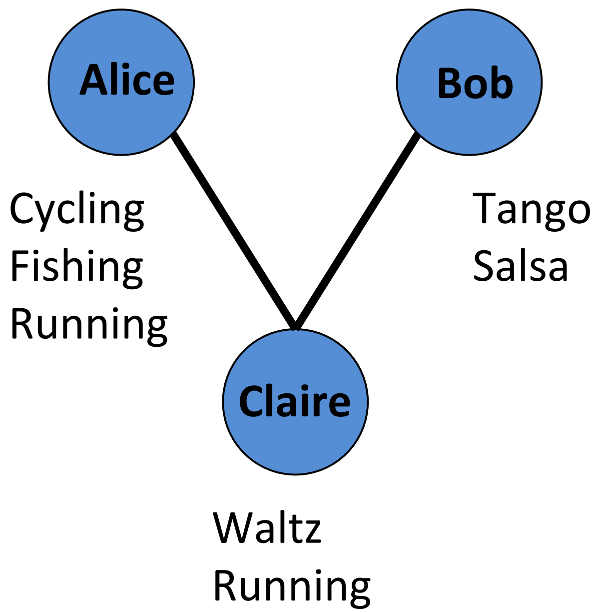
$W_{kk'}$: Probability that group k connects to group k'

For each pair of nodes (i, j)

$$z_i^{(ij)} \sim \mathrm{discrete}(\pi^{(i)})$$
$$z_j^{(ij)} \sim \mathrm{discrete}(\pi^{(j)})$$
$$Y_{ij} \sim \mathrm{Bernoulli}(W_{z_i^{(ij)}, z_j^{(ij)}})$$

Airoldi et al., (2008)

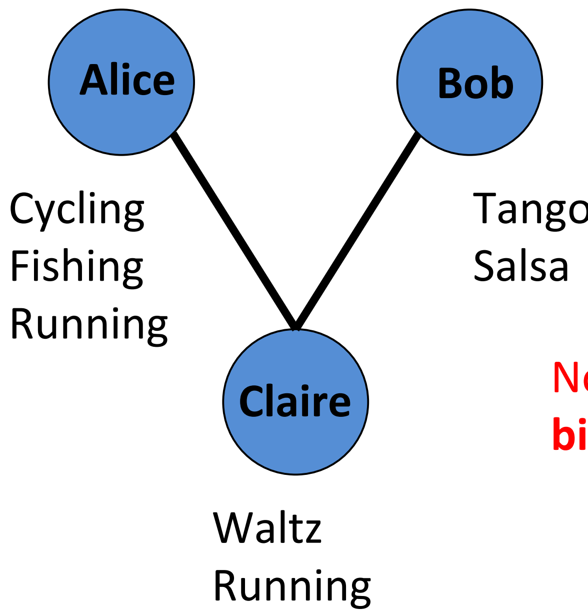
Latent feature models



Mixed membership implies a kind of “conservation of (probability) mass” constraint:
If you like cycling more, you must like running less, to sum to one

Miller, Griffiths, Jordan (2009)

Latent feature models



Nodes represented by
binary vector of latent features

Z =

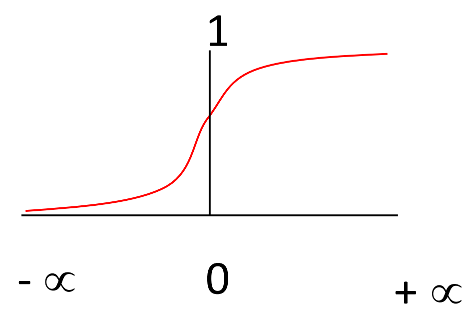
	Cycling	Fishing	Running	Tango	Salsa	Waltz
Alice						
Bob						
Claire						

Miller, Griffiths, Jordan (2009)

Latent feature models

- Latent Feature Relational Model LFRM (Miller, Griffiths, Jordan, 2009) likelihood model:

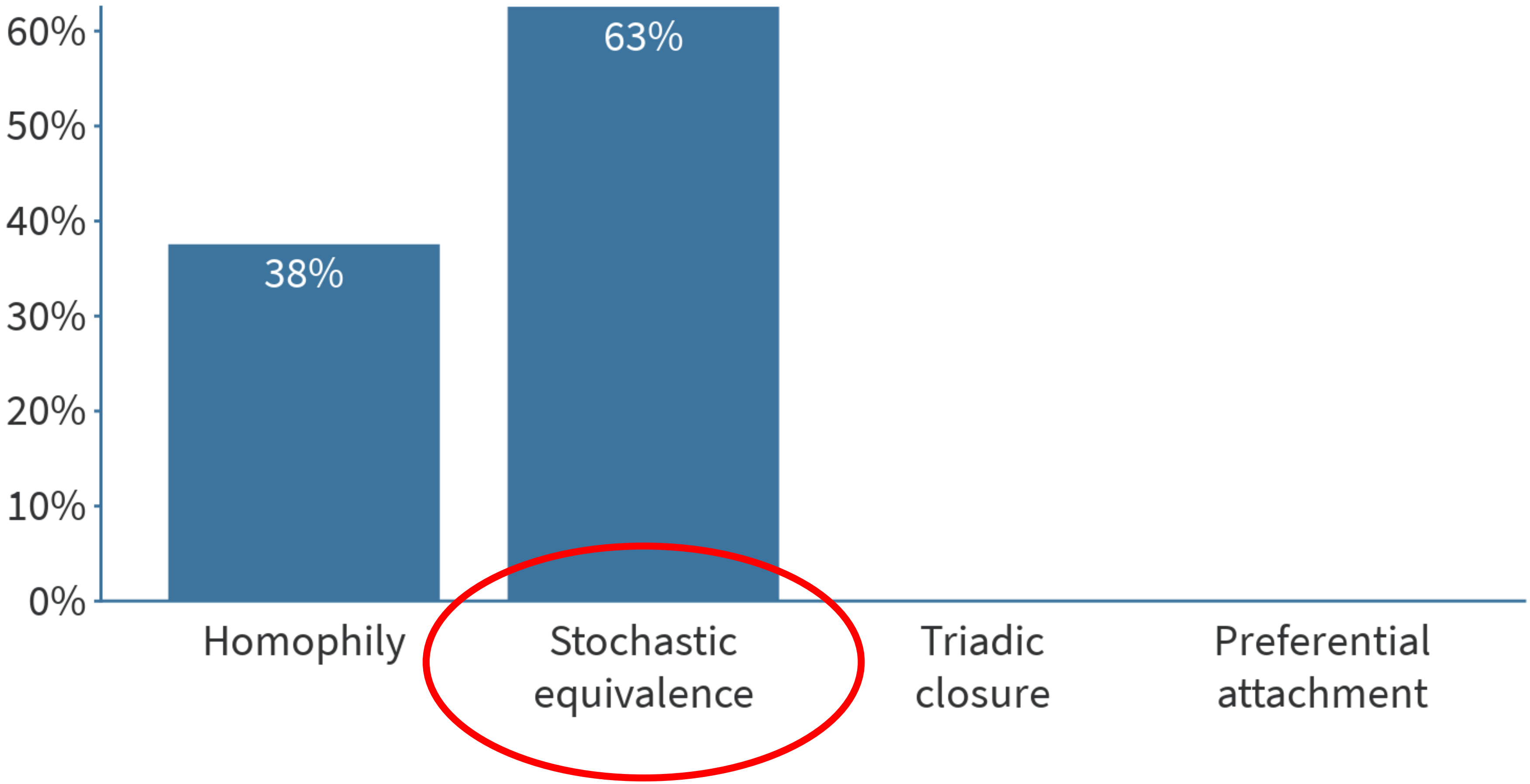
$$P(Y_{ij} = 1 | \ldots) = \sigma(\mathbf{z}_i \mathbf{W} \mathbf{z}_j^{\top})$$



- "If I have feature *k*, and you have feature *l*, add W_{kl} to the log-odds of the probability we interact"
- Can include terms for network density, covariates, popularity, etc.

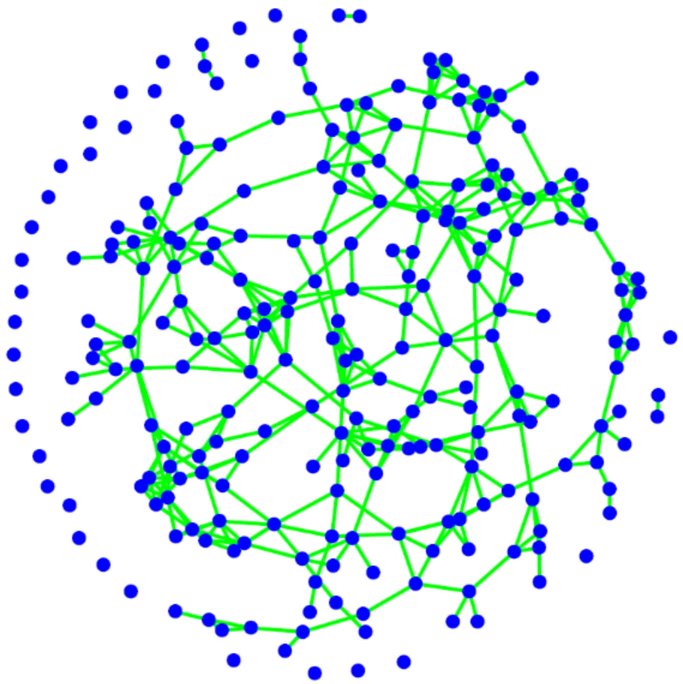
A block model is a model of network data that relies on the notion of _____

Poll locked. Responses not accepted.



Which is a more appropriate model for this dataset?

Respond at **PollEv.com/jamesfoulds656** Text **JAMESFOULDS656** to **37607** once to join, then **A or B**



Blockmodel **A**

Latent space model **B**

Total Results: 22

Python code for demos available on tutorial website

https://github.com/kevin-s-xu/ICWSM-2018-Generative-Tutorial

Outline

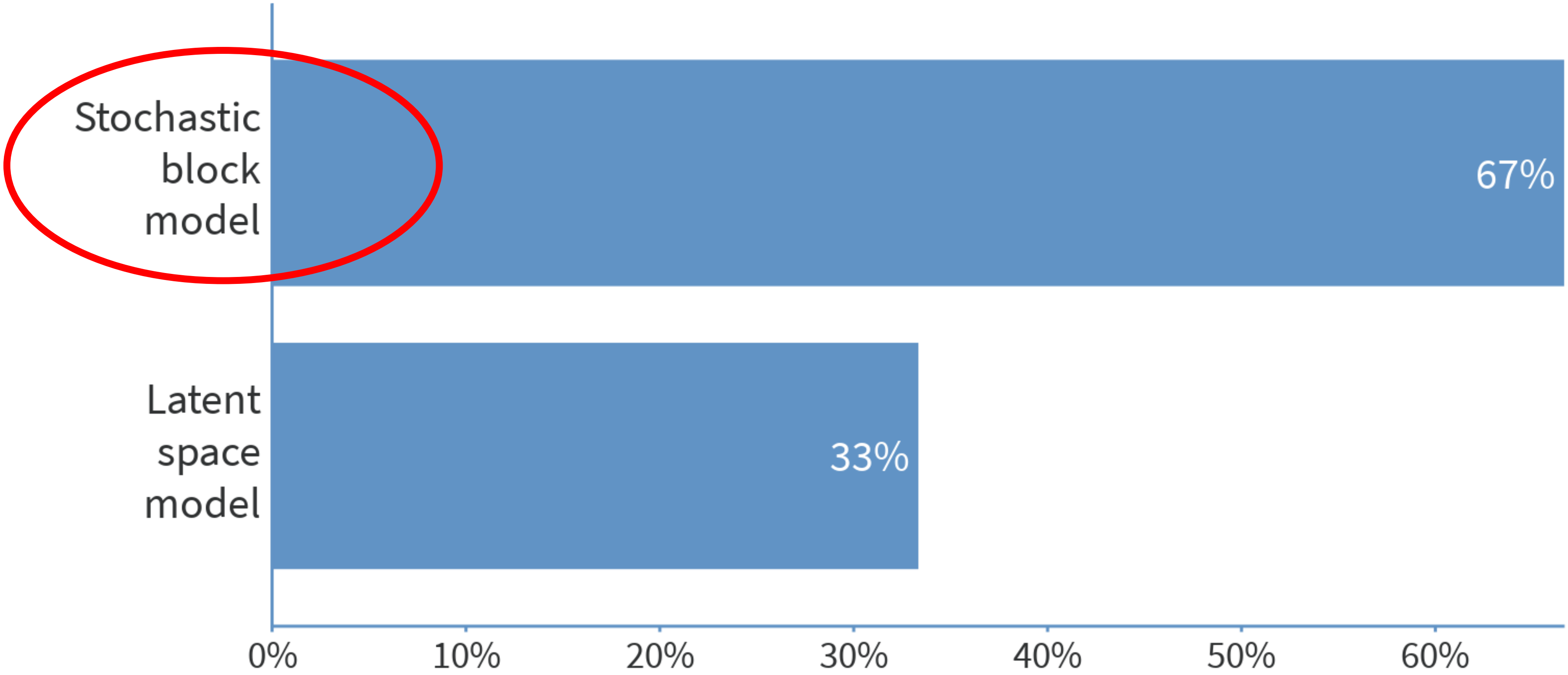
- Mathematical representations and generative models for social networks
  - Introduction to generative approach
  - Connections to sociological principles
- Fitting generative social network models to data
  - Application scenarios with demos
  - Model selection and evaluation
- Rich generative models for social media data
  - Network models augmented with text and dynamics
  - Case studies on social media data

Application 1: Facebook wall posts

- Network of wall posts on Facebook collected by Viswanath et al. (2009)
  - Nodes: Facebook users
  - Edges: directed edge from i to j if i posts on j’s Facebook wall
- What model should we use?

We have a network of wall posts on Facebook. Nodes: Facebook users. Edges: directed edge from i to j if i posts on j's Facebook wall. Which model should we use?

Poll locked. Responses not accepted.

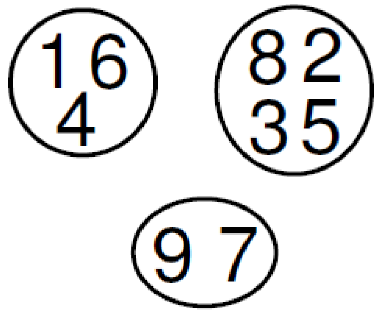


Application 1: Facebook wall posts

- Network of wall posts on Facebook collected by Viswanath et al. (2009)
  - Nodes: Facebook users
  - Edges: directed edge from i to j if i posts on j's Facebook wall
- What model should we use?
  - (Continuous) latent space models do not handle directed graphs in a straightforward manner
  - Wall posts might not be transitive, unlike friendships
- Stochastic block model might not be a bad choice as a starting point

Model structure

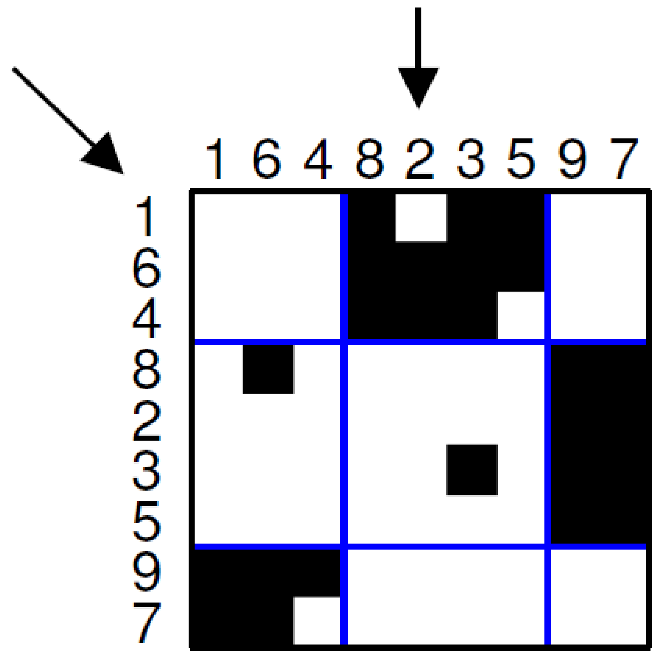
Latent groups Z



Interaction matrix W

0.1	0.9	0.1
0.1	0.1	0.9
0.9	0.1	0.1

(probability of an edge from block k to block k')



Kemp, Charles, et al. "Learning systems of concepts with an infinite relational model." AAAI. Vol. 3. 2006.

Fitting stochastic block model

- A priori block model: assume that class (role) of each node is given by some other variable
  - Only need to estimate $W_{kk'}$: probability that node in class k connects to node in class k' for all k, k'
- Likelihood given by

Number of actual edges in block (k, k')

Number of possible edges in block (k, k')

$$Pr(\mathbf{Y}|\mathbf{W}, \mathbf{Z})$$
$$= \exp\left\{\sum_{k=1}^{K}\sum_{k'=1}^{K}\left[m_{kk'}\log W_{kk'} + (n_{kk'} - m_{kk'})\log(1 - W_{kk'})\right]\right\}$$

- Maximum-likelihood estimate (MLE) given by

$$\hat{W}_{kk'} = \frac{m_{kk'}}{n_{kk'}}$$

Estimating latent classes

- Latent classes (roles) are unknown in this data set
  - First estimate latent classes $\mathbf{Z}$ then use MLE for $\mathbf{W}$
- MLE over latent classes is intractable!
  - $\sim K^N$ possible latent class vectors
- Spectral clustering techniques have been shown to accurately estimate latent classes
  - Use singular vectors of (possibly transformed) adjacency matrix to estimate classes
  - Many variants with differing theoretical guarantees

Spectral clustering for directed SBMs

1. Compute singular value decomposition $Y = U\Sigma V^T$
2. Retain only first K columns of U, V and first K rows and columns of Σ
3. Define coordinate-scaled singular vector matrix $\tilde{Z} = \left[U\Sigma^{1/2}\ V\Sigma^{1/2}\right]$
4. Run k-means clustering on rows of $\tilde{Z}$ to return estimate $\hat{Z}$ of latent classes

Scales to networks with thousands of nodes!

Demo of SBM on Facebook wall post network

1. Load adjacency matrix $\mathbf{Y}$
2. Model selection: examine singular values of $\mathbf{Y}$ to choose number of latent classes (blocks)
   - Eigengap heuristic: look for gaps between singular values
3. Fit selected model
4. Analyze model fit: class memberships and block-dependent edge probabilities
5. Simulate new networks from model fit
6. Check how well simulated networks preserve actual network properties (posterior predictive check)

Conclusions from posterior predictive check

- Block densities are well-replicated
- Transitivity is partially replicated
  - No mechanism for transitivity in SBM so this is a natural consequence of block-dependent edge probabilities
- Reciprocity is not replicated at all
  - Pair-dependent stochastic block model can be used to preserve reciprocity

$$p(\mathbf{Y}|\mathbf{Z},\theta) = \prod_{i \neq j} p\left(y_{ij}, y_{ji} \middle| \mathbf{z}_i, \mathbf{z}_j, \theta\right)$$

  - 4 choices for pair or dyad: $\left(y_{ij}, y_{ji}\right) \in \{(0,0),(0,1),(1,0),(1,1)\}$

Application 2: Facebook friendships

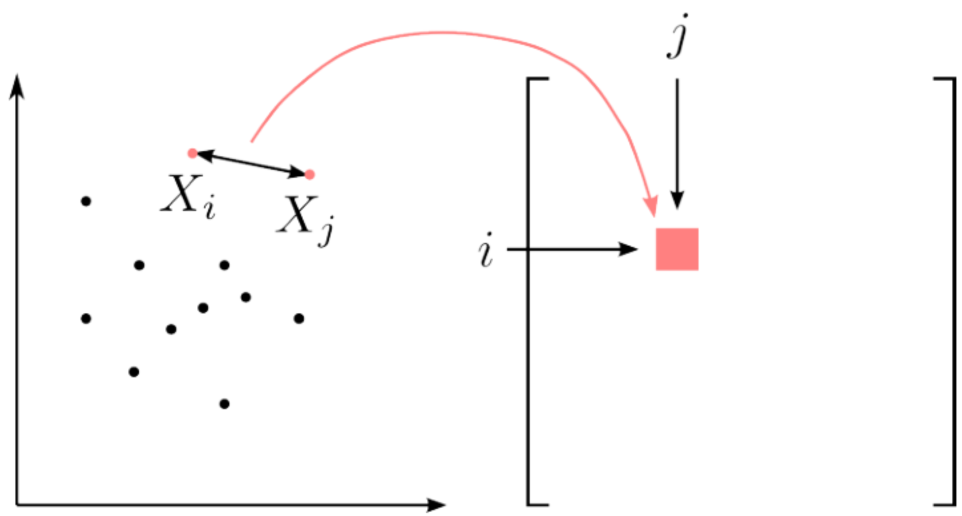
- Network of friendships on Facebook collected by Viswanath et al. (2009)
  - Nodes: Facebook users
  - Edges: undirected edge between i and j if they are friends
- What model should we use?

Application 2: Facebook friendships

- Network of friendships on Facebook collected by Viswanath et al. (2009)
  - Nodes: Facebook users
  - Edges: undirected edge between i and j if they are friends
- What model should we use?
  - Edges denote friendships so lots of transitivity may be expected (compared to wall posts)
  - Stochastic block model can replicate some transitivity due to class-dependent edge probabilities but doesn't explicitly model transitivity
- Latent space model might be a better choice

(Continuous) latent space model

- (Continuous) latent space model (LSM) proposed by Hoff et al. (2002)
  - Each node has a latent position $\mathbf{z}_i \in \mathbb{R}^d$
  - Probabilities of forming edges depend on distances between latent positions
  - Define pairwise affinities $\psi_{ij} = \theta - \left\|\mathbf{z}_i - \mathbf{z}_j\right\|_2$



$$p(\mathbf{Y}|\mathbf{Z}, \theta) = \prod_{i \neq j} \frac{e^{y_{ij}\psi_{ij}}}{1 + e^{\psi_{ij}}}$$

Estimation for latent space model

- Maximum-likelihood estimation
  - Log-likelihood is concave in terms of pairwise distance matrix D but not in latent positions Z
  - First find MLE in terms of D then use multi-dimensional scaling (MDS) to get initialization for Z
  - Faster approach: replace D with shortest path distances in graph then use MDS
  - Use quasi-Newton (BFGS) optimization to find MLE for Z
- Latent space dimension often set to 2 to allow visualization using scatter plot

Scales to ~1000 nodes

Demo of latent space model on Facebook friendship network

1. Load adjacency matrix $\mathbf{Y}$
2. Model selection: choose dimension of latent space
   - Typically start with 2 dimensions to enable visualization
3. Fit selected model
4. Analyze model fit: examine estimated positions of nodes in latent space and estimated bias
5. Simulate new networks from model fit
6. Check how well simulated networks preserve actual network properties (posterior predictive check)

Conclusions from posterior predictive check

- Block densities are well-replicated by SBM
- Transitivity is partially replicated by SBM
- Overall density is well-replicated by latent space model
    - No blocks in latent space model
- Transitivity is well-replicated by latent space model
- Can increase dimension of latent space if posterior check reveals poor fit
    - Not needed in this small network

Frequentist inference

- Both these demos used frequentist inference
- Parameters θ treated as having fixed but unknown values
  - Stochastic block model parameters: class memberships $\mathbf{Z}$ and block-dependent edge probabilities $\mathbf{W}$
  - Latent space model parameters: latent node positions $\mathbf{Z}$ and scalar global bias θ
- Estimate parameters by maximizing likelihood function of the parameters

$$\hat{\theta}_{MLE} = \mathrm{argmax}_\theta \, Pr(\mathbf{X}|\theta)$$

Bayesian inference

- Parameters θ treated as random variables. We can then take into account uncertainty over them

- As a Bayesian, all you have to do is write down your prior beliefs, write down your likelihood, and apply Bayes ' rule,

$$Pr(\theta|\mathbf{X}) = \frac{Pr(\mathbf{X}|\theta)Pr(\theta)}{Pr(\mathbf{X})}$$

Elements of Bayesian Inference

Likelihood

$$Pr(\theta|\mathbf{X}) = \frac{Pr(\mathbf{X}|\theta)Pr(\theta)}{Pr(\mathbf{X})}$$

Prior

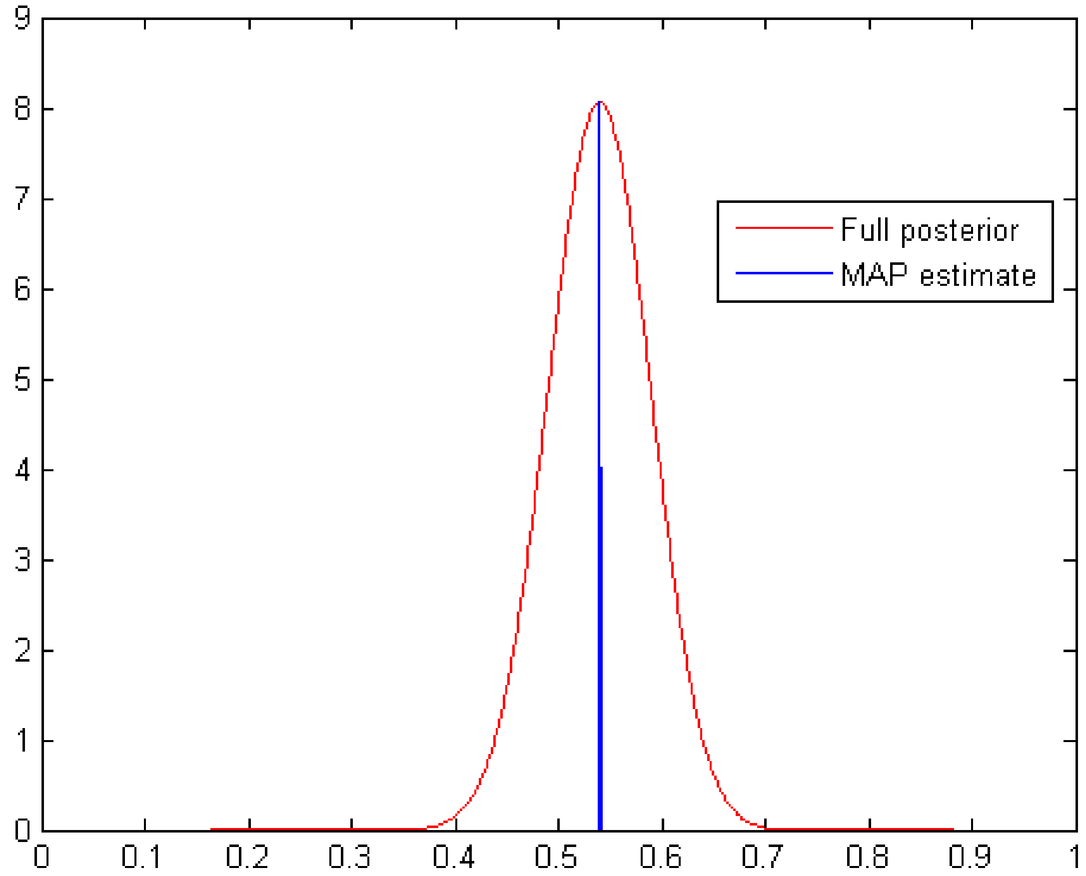
Posterior

Marginal likelihood
(a.k.a. model evidence)

$$Pr(\mathbf{X}) = \int Pr(\mathbf{X}|\theta)Pr(\theta)d\theta$$

is a normalization constant that does not depend on the value of θ. It is the probability of the data under the model, marginalizing over all possible θ's.

MAP estimate can result in overfitting



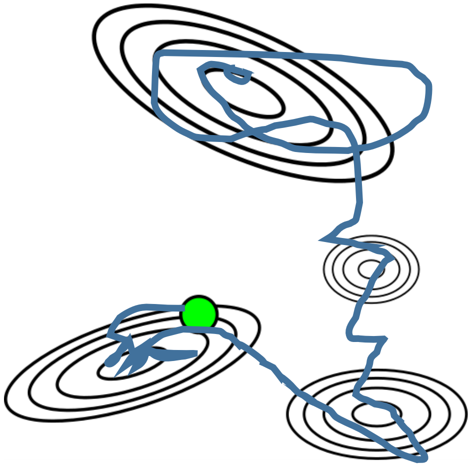
Inference Algorithms

- **Exact inference**
  - Generally intractable
- **Approximate inference**
  - **Optimization approaches**
    - EM, variational inference
  - **Simulation approaches**
    - Markov chain Monte Carlo, importance sampling, particle filtering

Markov chain Monte Carlo

- **Goal**: approximate/summarize a distribution, e.g. the posterior, with a set of samples

- **Idea**: use a Markov chain to simulate the distribution and draw samples



Gibbs sampling

- Update variables one at a time by drawing from their conditional distributions

$$\mathbf{z}_i := \mathbf{z}_i^{(new)},\ \mathbf{z}_i^{(new)} \sim Pr(\mathbf{z}_i | \mathbf{z}_{\neg i})$$

- In each iteration, sweep through and update all of the variables, in any order.

Gibbs sampling for SBM

Initialize group assignments and parameters randomly

Until converged

For each pair of groups k, k'

$$W_{kk'} \sim \mathrm{Beta}(n^{(1)}_{kk'} + \alpha_1, n^{(0)}_{kk'} + \alpha_0)$$

$$\pi \sim \mathrm{Dirichlet}([n_1 + \alpha_1, \ldots, n_K + \alpha_K])$$

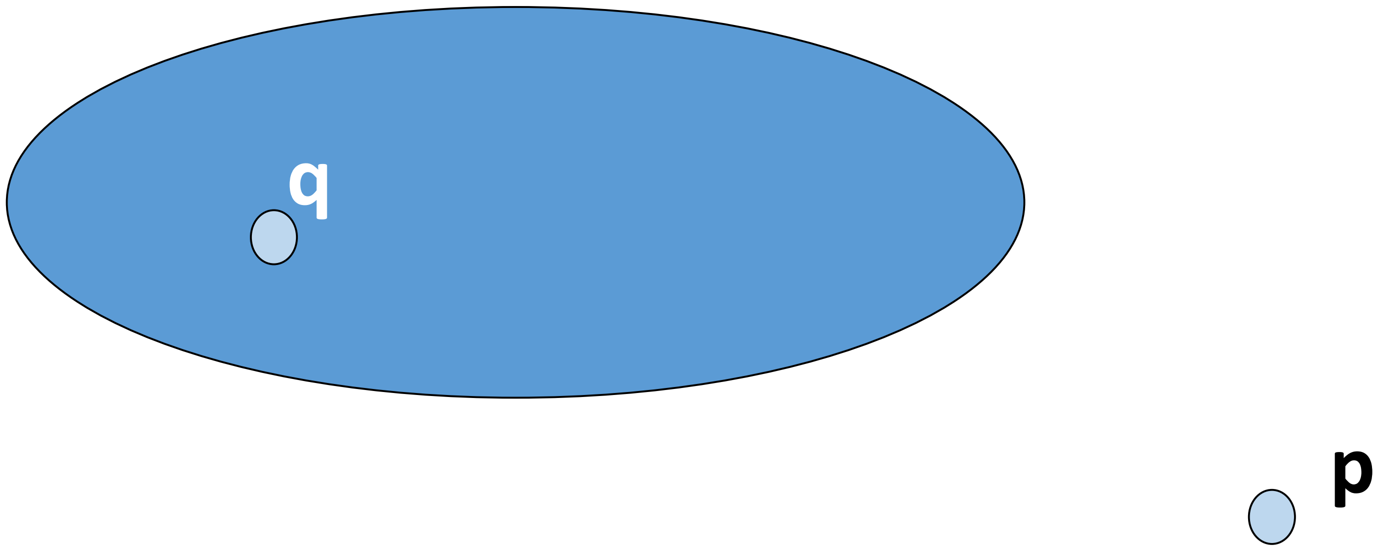
For node i

$$Pr(z_i = k) \propto \pi_k \prod_{k'=1}^{K} W_{kk'}^{n^{(1)}_{i,k'}} (1 - W_{kk'})^{n^{(0)}_{i,k'}}$$

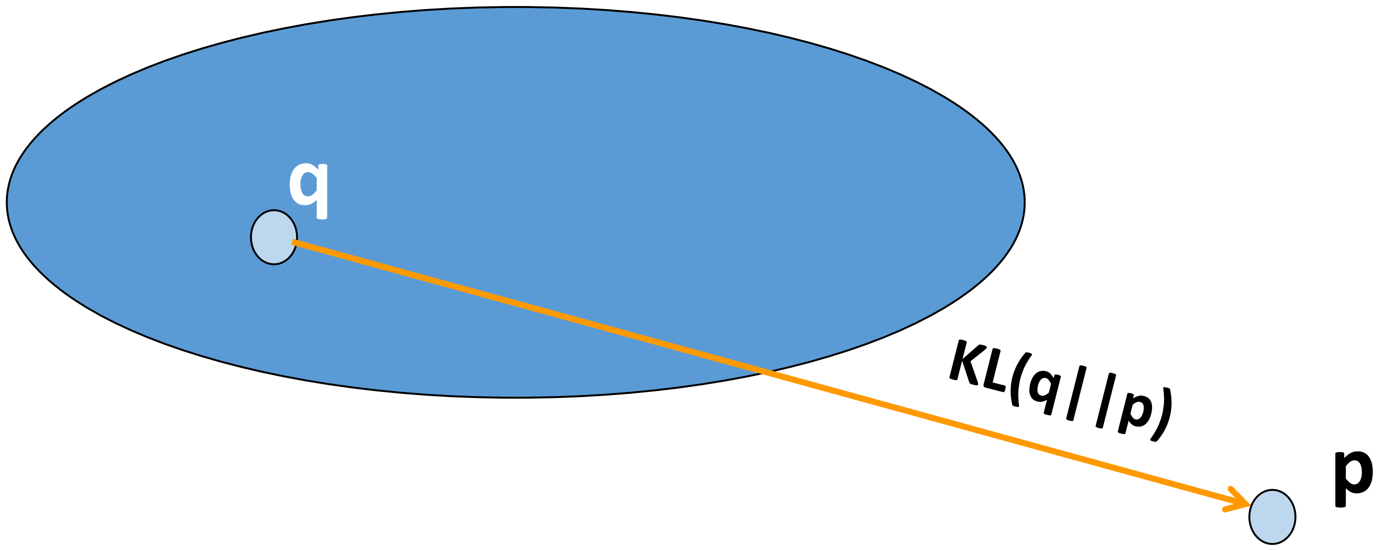
Variational inference

- Key idea:
  - Approximate distribution of interest p(z) with another distribution q(z)
  - Make q(z) tractable to work with
  - Solve an optimization problem to make q(z) as similar to p(z) as possible, e.g. in KL-divergence

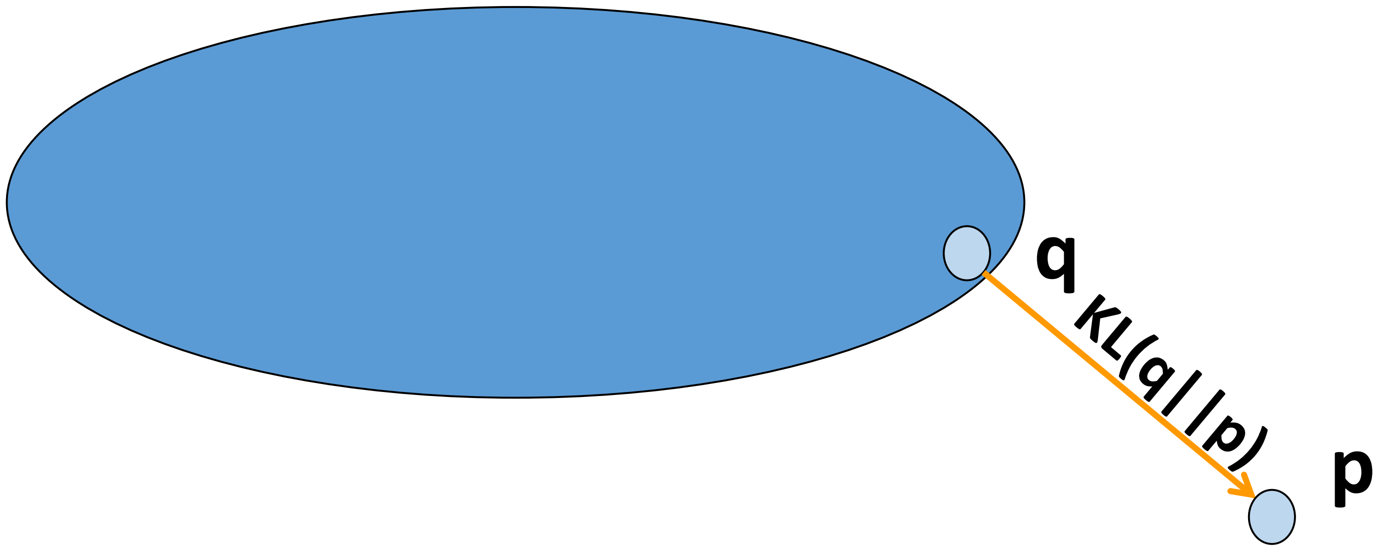
Variational inference



Variational inference



Variational inference



Mean field algorithm

- The mean field approach uses a fully factorized *q(**z**)*

$$q(\mathbf{z}) = \prod_i q_i(z_i)$$

- Until converged
  - For each factor *i*
    - Select variational parameters γ_i such that

$$q_i(z_i|\gamma_i) :\propto \exp(E_{q_{\neg i}}[\log p(\mathbf{z}, \mathbf{x})])$$

Mean field vs Gibbs sampling

- Both mean field and Gibbs sampling iteratively update one variable given the rest

- Mean field stores an entire distribution for each variable, while Gibbs sampling draws from one.

Pros and cons vs Gibbs sampling

- **Pros:**
  - Deterministic algorithm, typically converges faster
  - Stores an analytic representation of the distribution, not just samples
  - Non-approximate parallel algorithms
  - Stochastic algorithms can scale to very large data sets
  - No issues with checking convergence

- **Cons:**
  - Will never converge to the true distribution, unlike Gibbs sampling
  - Dense representation can mean more communication for parallel algorithms
  - Harder to derive update equations

Variational inference algorithm for MMSB (Variational EM)

- Compute maximum likelihood estimates for interaction parameters $W_{kk'}$
- Assume fully factorized variational distribution for mixed membership vectors, cluster assignments

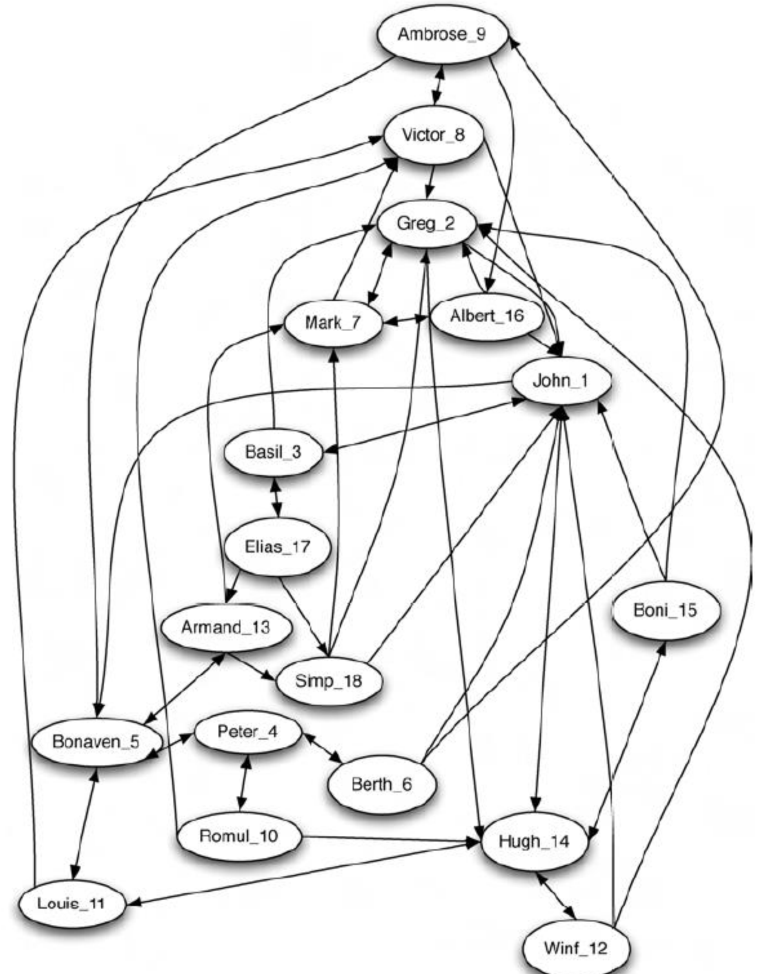
- Until converged
  - For each node
    - Compute variational discrete distribution over it's latent $z_{p\text{->}q}$ and $z_{q\text{->}p}$ assignments
    - Compute variational Dirichlet distribution over its mixed membership distribution
  - Maximum likelihood update for $\boldsymbol{W}$

Application of MMSB to Sampson's Monastery

- Sampson (1968) studied friendship relationships between novice monks

- Identified several factions
  - Blockmodel appropriate?

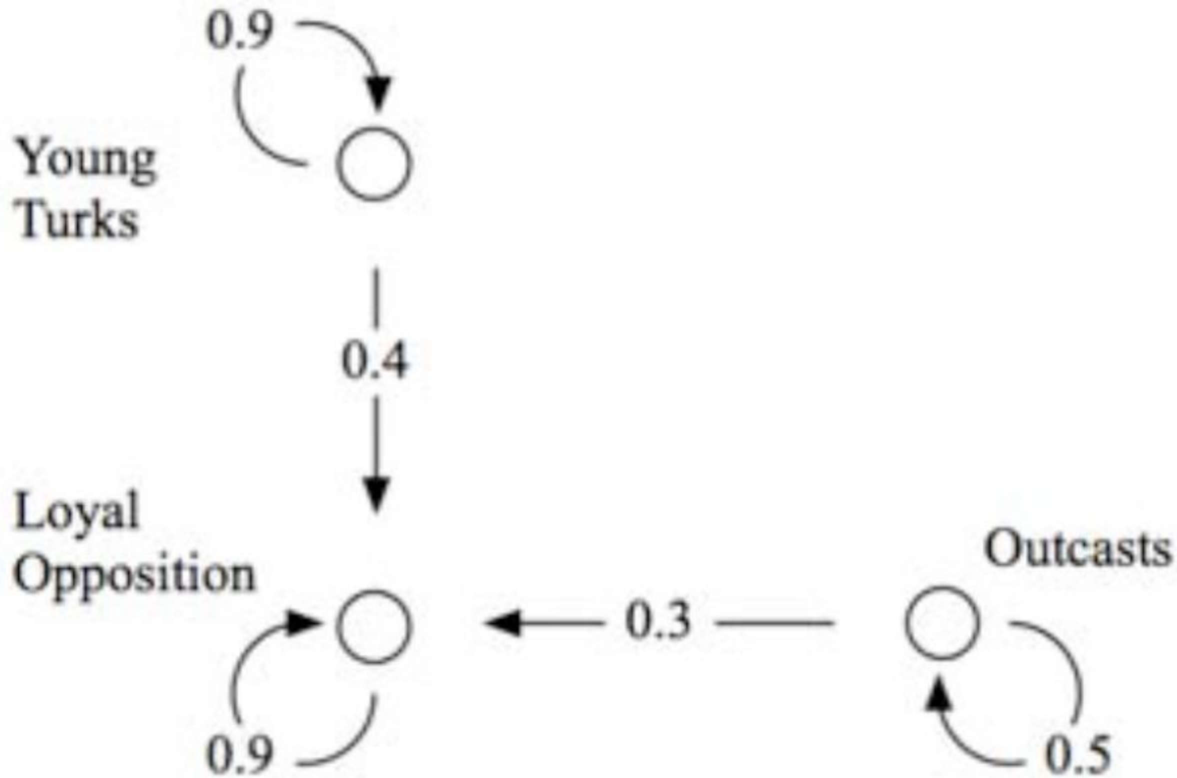
- Conflicts occurred
  - Two monks expelled
  - Others left



Airoldi, E. M., Blei, D. M., Fienberg, S. E., & Xing, E. P. (2009). Mixed membership stochastic blockmodels. In Advances in Neural Information Processing Systems (pp. 33-40).

Application of MMSB to Sampson's Monastery

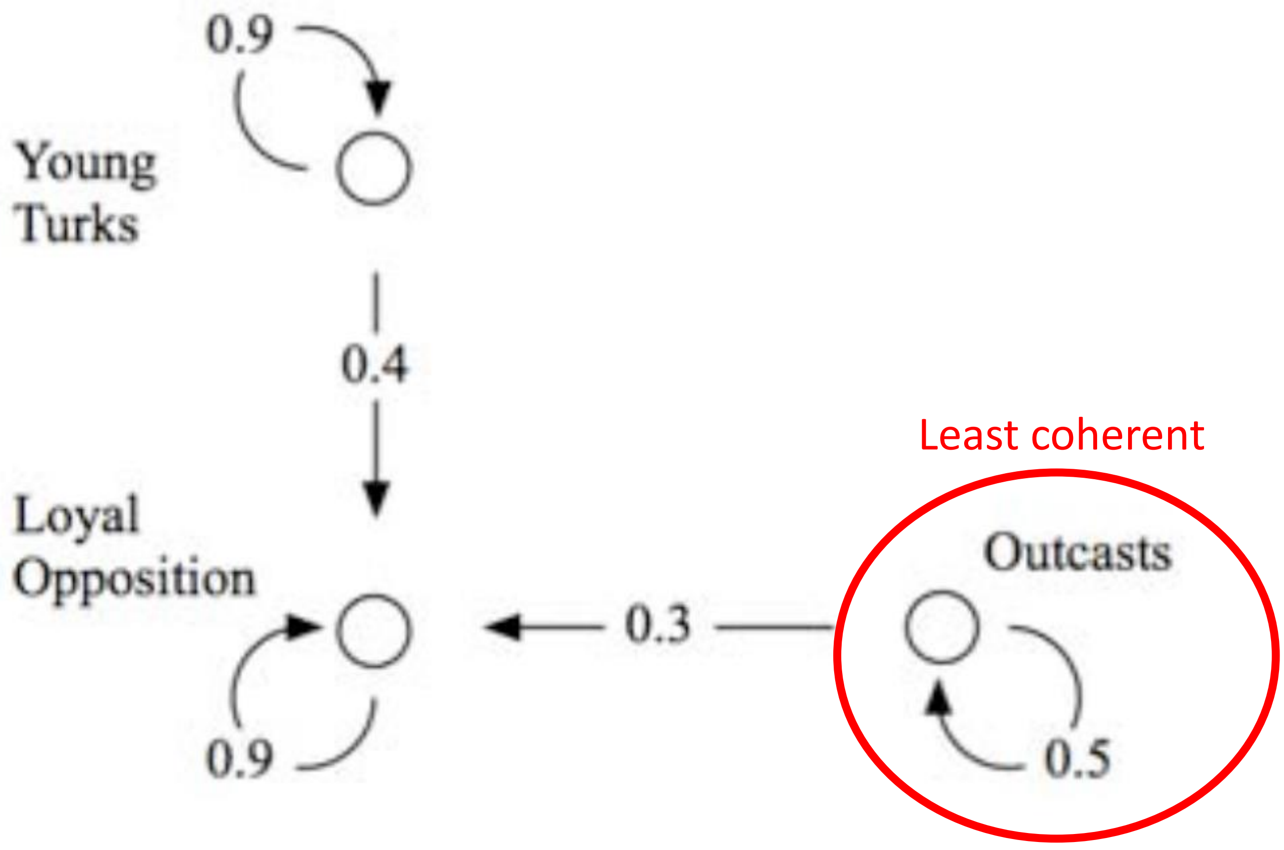
Estimated blockmodel



Airoldi, E. M., Blei, D. M., Fienberg, S. E., & Xing, E. P. (2009). Mixed membership stochastic blockmodels. In Advances in Neural Information Processing Systems (pp. 33-40).

Application of MMSB to Sampson's Monastery

Estimated blockmodel

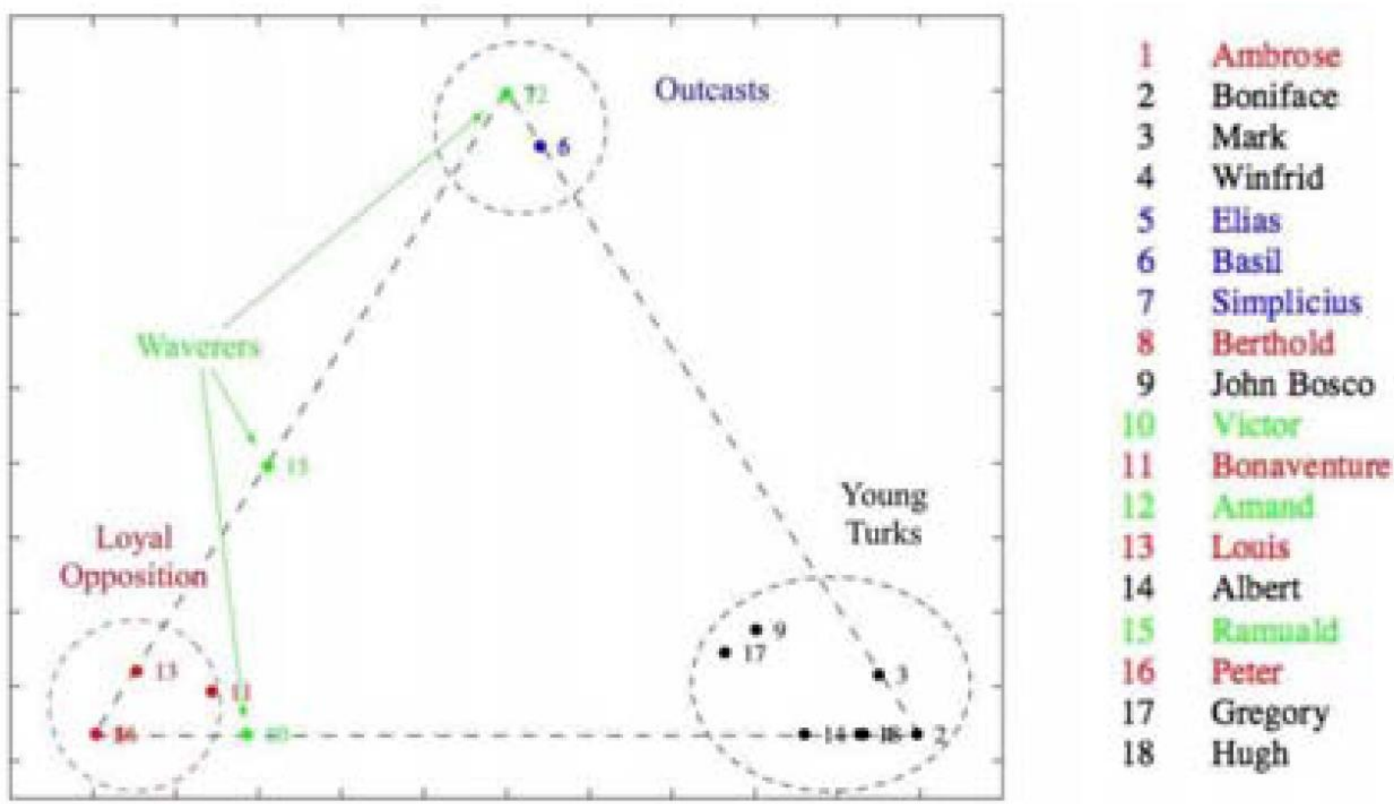


Airoldi, E. M., Blei, D. M., Fienberg, S. E., & Xing, E. P. (2009). Mixed membership stochastic blockmodels. In Advances in Neural Information Processing Systems (pp. 33-40).

Application of MMSB to Sampson's Monastery

Estimated Mixed membership vectors

(posterior mean)

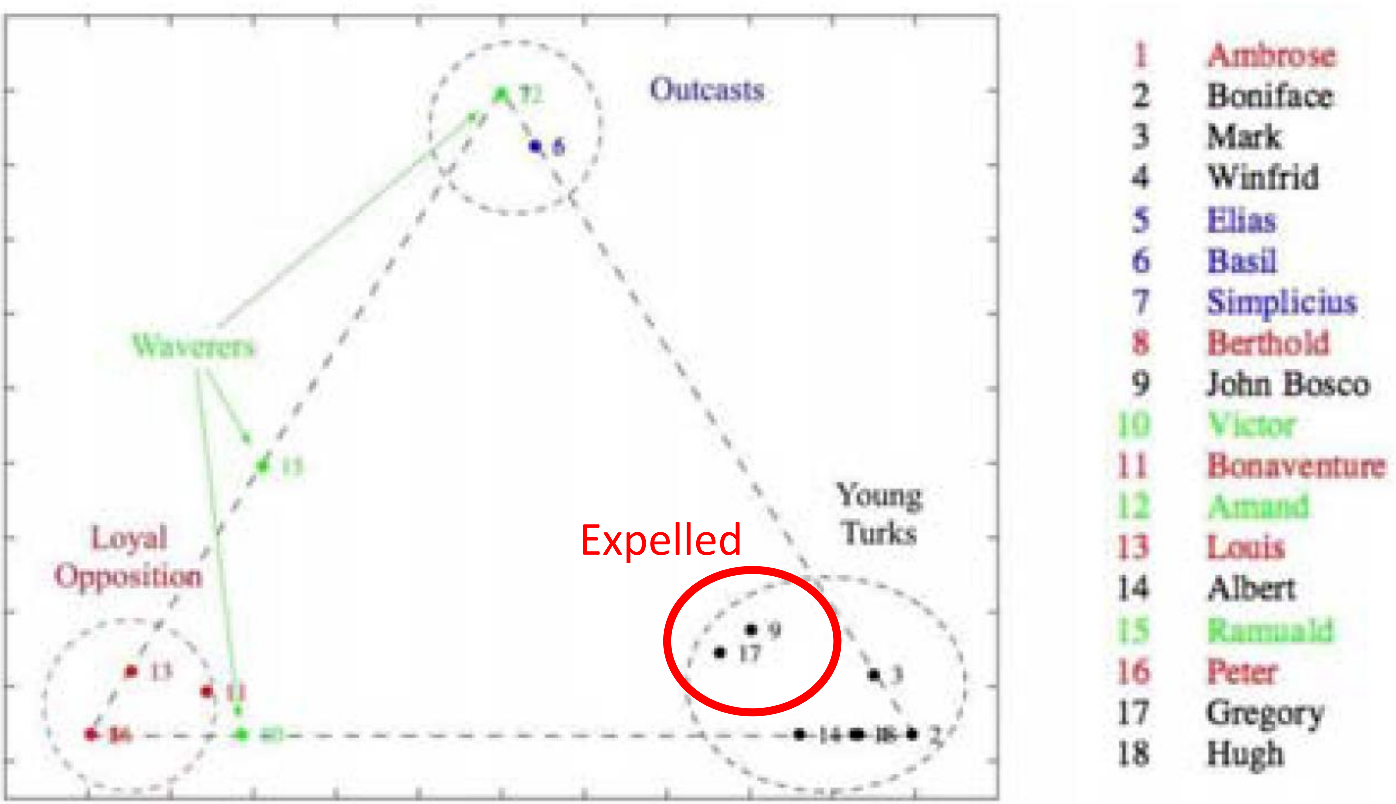


Airoldi, E. M., Blei, D. M., Fienberg, S. E., & Xing, E. P. (2009). Mixed membership stochastic blockmodels. In Advances in Neural Information Processing Systems (pp. 33-40).

Application of MMSB to Sampson's Monastery

Estimated Mixed membership vectors

(posterior mean)

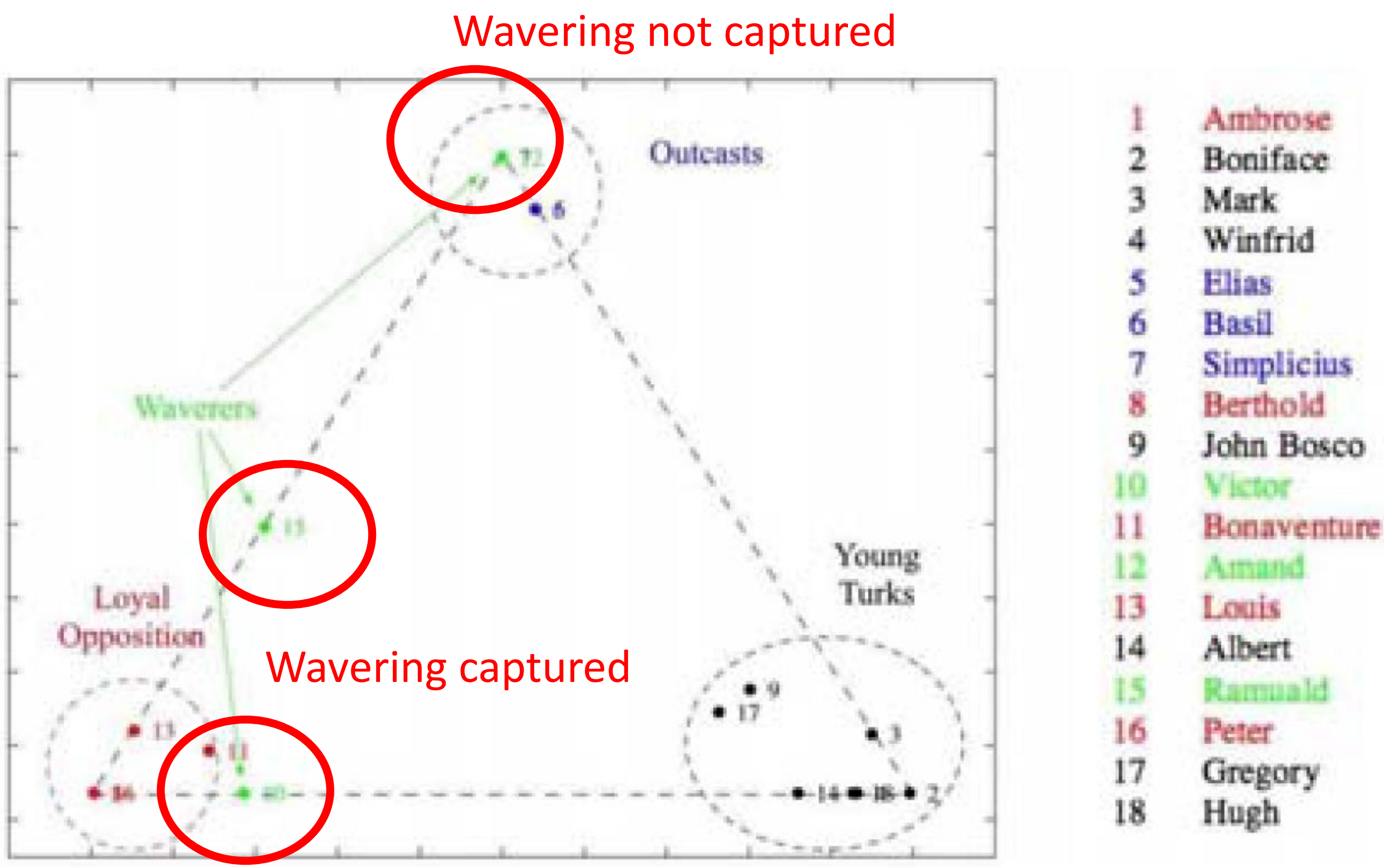


Airoldi, E. M., Blei, D. M., Fienberg, S. E., & Xing, E. P. (2009). Mixed membership stochastic blockmodels. In Advances in Neural Information Processing Systems (pp. 33-40).

Application of MMSB to Sampson's Monastery

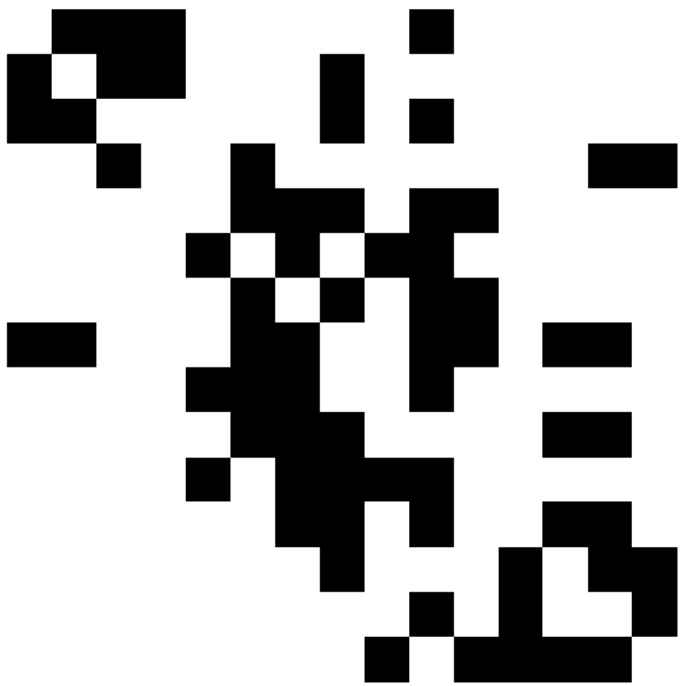
Estimated Mixed membership vectors

(posterior mean)

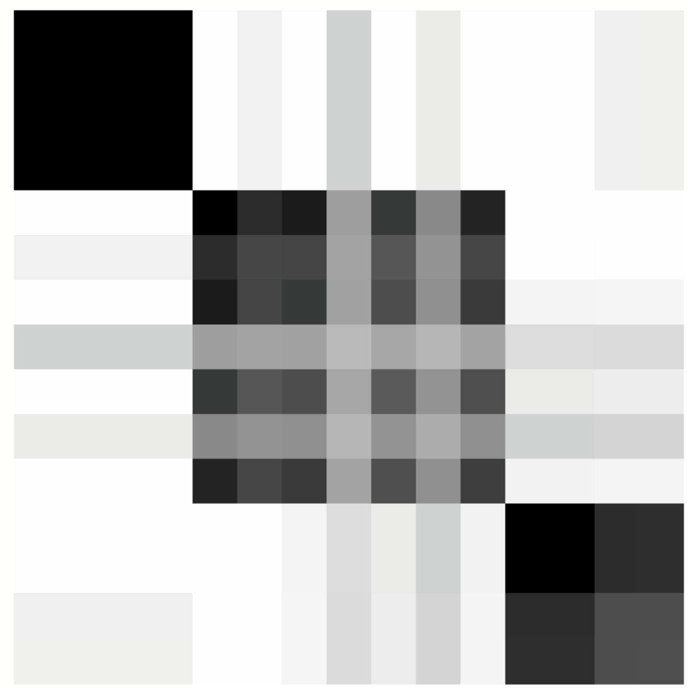


Airoldi, E. M., Blei, D. M., Fienberg, S. E., & Xing, E. P. (2009). Mixed membership stochastic blockmodels. In Advances in Neural Information Processing Systems (pp. 33-40).

Application of MMSB to Sampson's Monastery



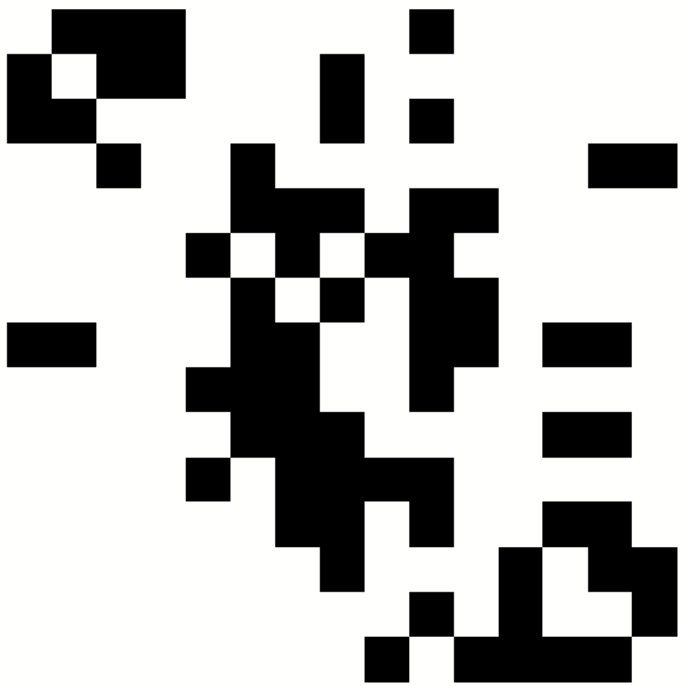
Original network
(whom do you like?)



Summary of network (use π's)

Airoldi, E. M., Blei, D. M., Fienberg, S. E., & Xing, E. P. (2009). Mixed membership stochastic blockmodels. In Advances in Neural Information Processing Systems (pp. 33-40).

Application of MMSB to Sampson's Monastery



Original network
(whom do you like?)

Denoise network (use z's)

Airoldi, E. M., Blei, D. M., Fienberg, S. E., & Xing, E. P. (2009). Mixed membership stochastic blockmodels. In Advances in Neural Information Processing Systems (pp. 33-40).

Evaluation of unsupervised models

- **Quantitative** evaluation
  - Measurable, quantifiable performance metrics

- **Qualitative** evaluation
  - Exploratory data analysis (EDA) using the model
  - Human evaluation, user studies,…

Evaluation of unsupervised models

- **Intrinsic** evaluation
  - Measure inherently good properties of the model
    - Fit to the data (e.g. link prediction), interpretability,…
- **Extrinsic** evaluation
  - Study usefulness of model for external tasks
    - Classification, retrieval, part of speech tagging,…

Extrinsic evaluation: What will you use your model for?

- If you have a **downstream task** in mind, you should probably evaluate based on it!

- Even if you don't, you could contrive one for evaluation purposes

- E.g. use latent representations for:
  - Classification, regression, retrieval, ranking…

Posterior predictive checks

- Sampling data from the posterior predictive distribution allows us to “look into the mind of the model” – G. Hinton

“This use of the word *mind* is not intended to be metaphorical. We believe that a mental state is the state of a hypothetical, external world in which a high-level internal representation would constitute veridical perception. That hypothetical world is what the figure shows.” **Geoff Hinton et al. (2006), A Fast Learning Algorithm for Deep Belief Nets.**

Posterior predictive checks

- Does data drawn from the model differ from the observed data, in ways that we care about?

- PPC:
  - Define a discrepancy function (a.k.a. test statistic) T(**X**).
    - Like a test statistic for a p-value. How extreme is my data set?
  - Simulate new data $\mathbf{X}^{(rep)}$ from the posterior predictive
    - Use MCMC to sample parameters from posterior, then simulate data
  - Compute $T(\mathbf{X}^{(rep)})$ and T(**X**), compare. Repeat, to estimate:

$$\mathrm{PPC} = P(T(\mathbf{X}^{(rep)}) > T(\mathbf{X})|\mathbf{X})$$

Outline

- Mathematical representations and generative models for social networks
  - Introduction to generative approach
  - Connections to sociological principles
- Fitting generative social network models to data
  - Application scenarios with demos
  - Model selection and evaluation
- Rich generative models for social media data
  - Network models augmented with text and dynamics
  - Case studies on social media data

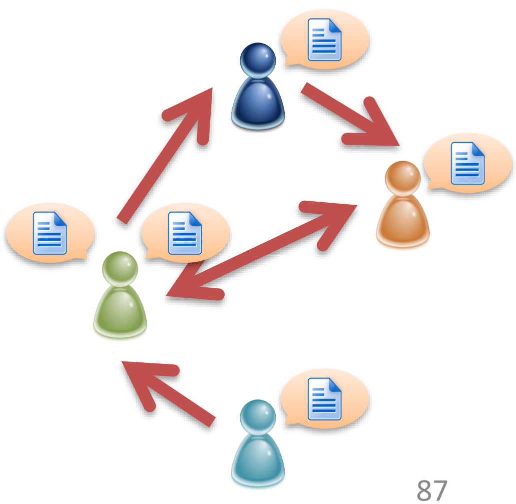
Networks and Text

- Social media data often involve **networks with text** associated
  - Tweets, posts, direct messages/emails,…

- Leveraging text can help to **improve network modeling**, and to **interpret** the network

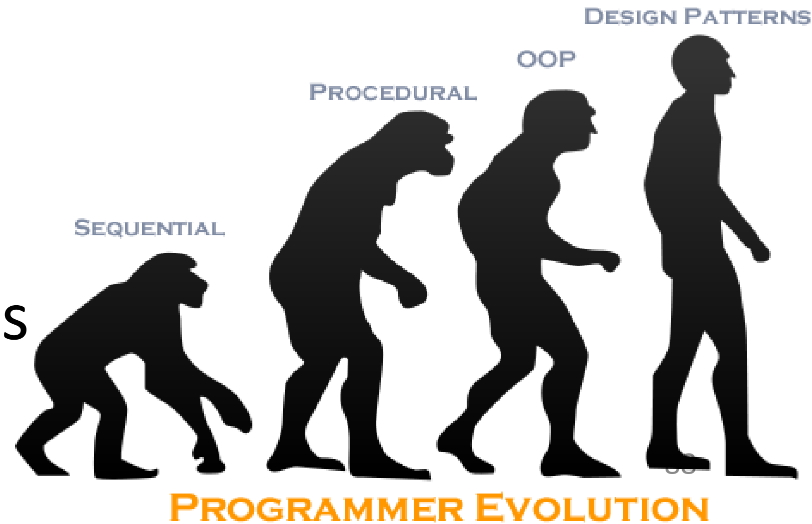
- Simple approach: model networks and text **separately**
  - Network model, can determine input for text analysis, e.g. the text for each network community

- More powerful methodology: **joint models of networks and text**
  - Usually combine network and language model components into a single model

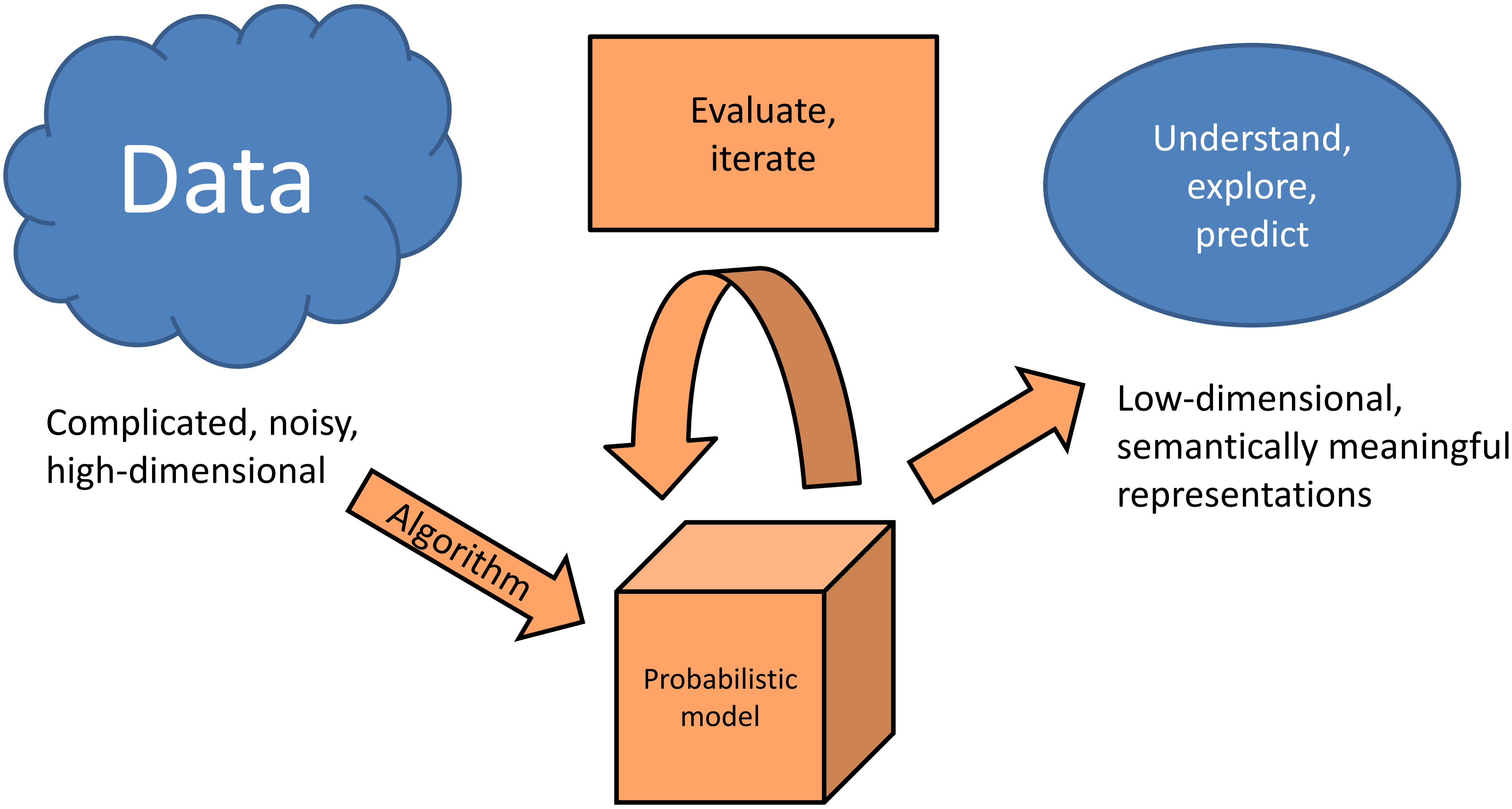


Design Patterns for Probabilistic Models

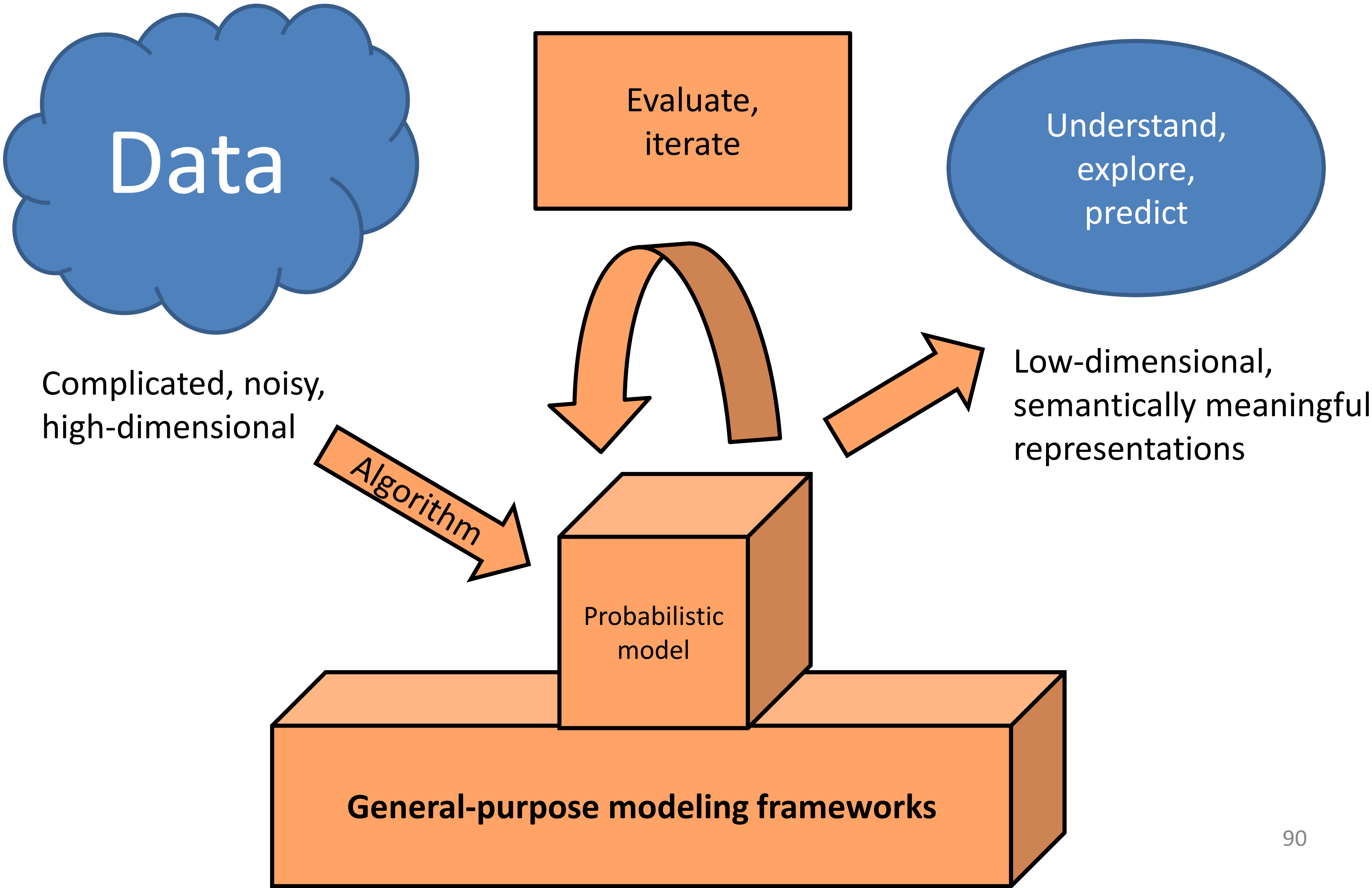
- **Condition** on useful information you **don't need to model**
- Or, **jointly model** multiple data modalities
- **Hierarchical/multi-level structure**
  - **Words in a document**
- **Graphical** dependencies
- **Temporal** modeling / time series



Box's Loop


Data
Complicated, noisy,
high-dimensional
Algorithm
Evaluate,
iterate
Probabilistic
model
Understand,
explore,
predict
Low-dimensional,
semantically meaningful
representations

Box's Loop


Data
Evaluate,
iterate
Understand,
explore,
predict
Complicated, noisy,
high-dimensional
Algorithm
Low-dimensional,
semantically meaningful
representations
Probabilistic
model
General-purpose modeling frameworks

Probabilistic Programming Languages

- These systems can make it **much easier for you** to develop **custom models** for social media analytics!

- Define a probabilistic model by writing code in a programming language

- The system automatically performs inference
  - Recently, these systems have become very practical

- Some popular languages:
  - Stan, Winbugs, JAGS, Infer.net, PyMC3, Edward, PSL

Infer.NET

- Imperative probabilistic programming API for any .NET language
- Multiple inference algorithms

Infer.NET user guide

A simple example

Here is an example of using Infer.NET to work out the probability of getting both heads when tossing two fair coins.

```
Variable<bool> firstCoin = Variable.Bernoulli(0.5);
Variable<bool> secondCoin = Variable.Bernoulli(0.5);
Variable<bool> bothHeads = firstCoin & secondCoin;
InferenceEngine ie = new InferenceEngine();
Console.WriteLine("Probability both coins are heads: "+ie.Infer(bothHeads));
```

The output of this program is:

```
Probability both coins are heads: Bernoulli(0.25)
```

Networked Frame Contests within #BlackLivesMatter Discourse

Drawing the Lines of Contention: Networked Frame Contests Within #BlackLivesMatter Discourse

LEO G. STEWART, Human Centered Design & Engineering, University of Washington
AHMER ARIF, Human Centered Design & Engineering, University of Washington
A. CONRAD NIED, Computer Science and Engineering, University of Washington
EMMA S. SPIRO, Information School and Sociology, University of Washington
KATE STARBIRD, Human Centered Design & Engineering, University of Washington

- Studies discourse around the **#BlackLivesMatter** movement on Twitter
- Finds **network communities** on the political left and right, and analyzes their competition in **framing** the issue
- The authors use a **mixed-method**, **interpretative** approach
  - Combination of **algorithms** and **qualitative content** analysis
  - Networks and text considered separately
    - network communities the focal points for qualitative study of text

Stewart et al. (2017). Drawing the Lines of Contention: Networked Frame Contests Within #BlackLivesMatter Discourse

Networked Frame Contests within #BlackLivesMatter Discourse

- **Retrieve tweets** using Twitter streaming API
  - between December 2015 – October 2016
  - keywords relating to both shootings and one of: blacklivesmatter, bluelivesmatter, alllivesmatter

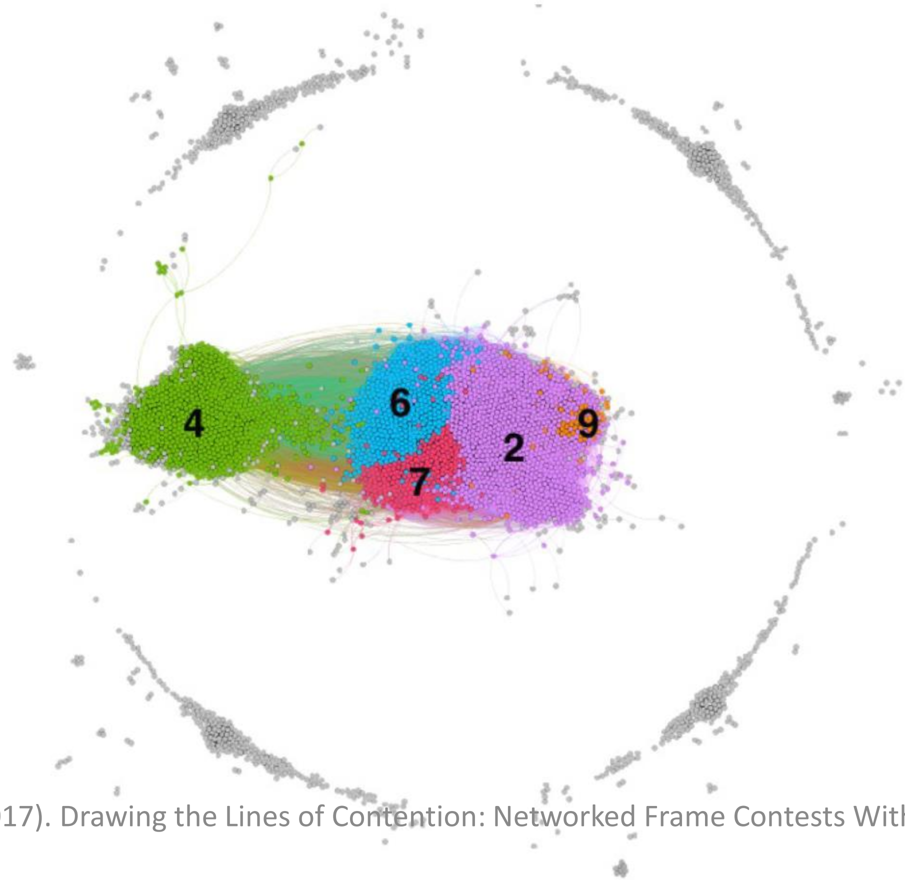
- Construct "**shared audience graph**"
  - Edges between users with large overlap in followers (20th percentile in Jaccard similarity of followers)

$$J(A,B) = \frac{|A_{followers}| \cap |B_{followers}|}{|A_{followers}| \cup |B_{followers}|}$$

Stewart et al. (2017). Drawing the Lines of Contention: Networked Frame Contests Within #BlackLivesMatter Discourse

Networked Frame Contests within #BlackLivesMatter Discourse

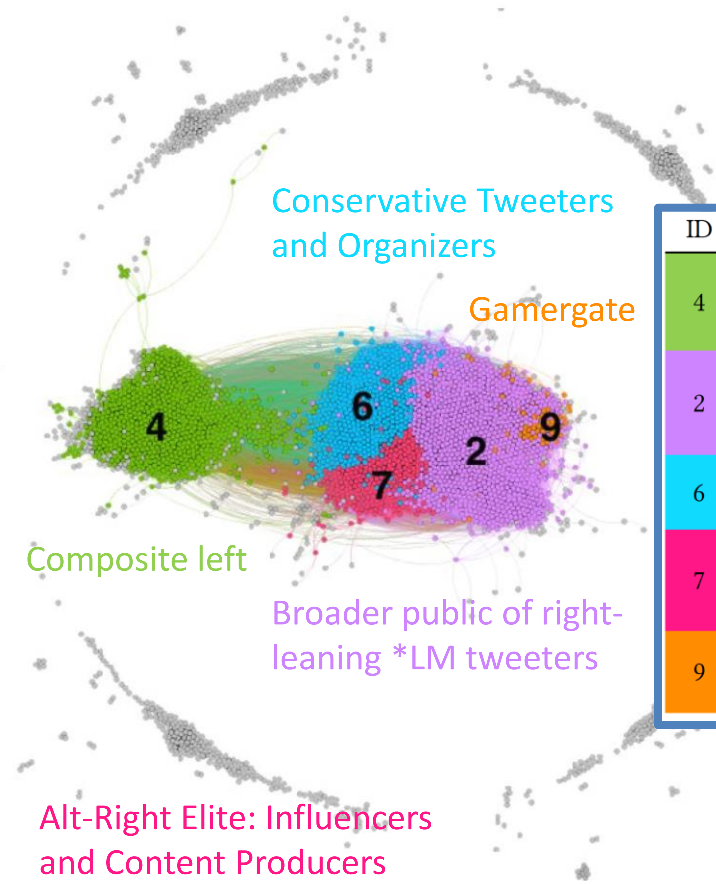
- Perform clustering on network to find communities
  - **Louvain modularity** method used. Aims to find densely connected clusters/communities with few connections to other communities



Stewart et al. (2017). Drawing the Lines of Contention: Networked Frame Contests Within #BlackLivesMatter Discourse

Networked Frame Contests within #BlackLivesMatter Discourse

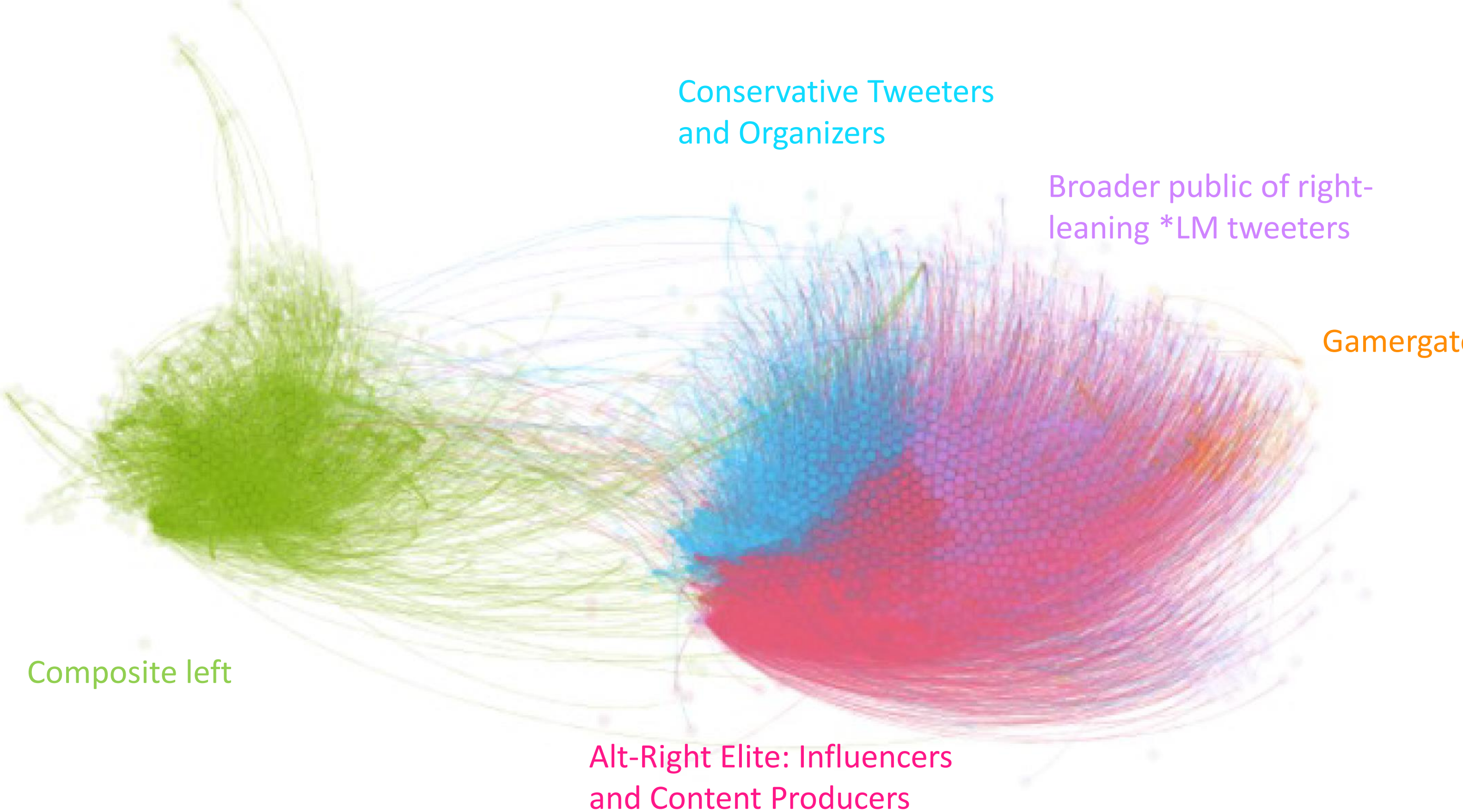
- Content analysis of the clusters



ID	Top 10 Hashtags in Account Profile Descriptions
4	#blacklivesmatter (262), #uniteblue (70), #imwithher (65), #feelthebern (50), #blm (45), #bernieorbust (39), #neverhillary (36), #nevertrump (32), #ourrevolution (25), #freepalestine (22)
2	#maga (157), #trump2016 (151), #2a (99), #bluelivesmatter (60), #trump (54), #neverhillary (53), #trumptrain (52), #makeamericagreatagain (50), #alllivesmatter (41), #nra (40)
6	#2a (146), #tcot (114), #nra (77), #pjnet (77), #maga (65), #trump2016 (62), #ccot (54), #neverhillary (50), #conservative (48), #bluelivesmatter (47)
7	#trump2016 (175), #maga (154), #2a (104), #trump (82), #makeamericagreatagain (71), #neverhillary (53), #americafirst (49), #trumptrain (49), #1a (44), #bluelivesmatter (44)
9	#gamergate (20), #opskynet (4), #maga (3), #notyourshield (3), #altright (2), #atheist (2), #blacklivesmatter (2), #gamer (2), #bringbackhanging (1), #teamvalor (1)

Stewart et al. (2017). Drawing the Lines of Contention: Networked Frame Contests Within #BlackLivesMatter Discourse

Networked Frame Contests within #BlackLivesMatter Discourse



Retweet Trajectories on the Shared Audience Graph

Very few retweets between left and right super-clusters (204/18,414 = 1.11%)

Stewart et al. (2017). Drawing the Lines of Contention: Networked Frame Contests Within #BlackLivesMatter Discourse

Networked Frame Contests within #BlackLivesMatter Discourse

- Study **framing contests** between left- and right-leaning super-clusters
- #BLM framing on the left: **injustice frames**

(Tweet 1): Cops called elderly Black man the n-word before shooting him to death #KillerCops #BlackLivesMatter

(Tweet 2): Recent acquittals of multiple officers involved in shootings makes Economic Boycott perfect for #BlackLivesMatter

(Tweet 3): Anyone blaming this Dallas shooting on the #BlackLivesMatter movement is sick. Those protestors were peaceful. This terrorized them too.

Stewart et al. (2017). Drawing the Lines of Contention: Networked Frame Contests Within #BlackLivesMatter Discourse

Networked Frame Contests within #BlackLivesMatter Discourse

- Study **framing contests** between left- and right-leaning super-clusters
- #BLM framing on the right: Reframing as **detrimental to social order** and being **anti-law**

(Tweet 4): Nothing Says #BlackLivesMatter like mass looting convenience stores & shooting ppl over the death of an armed thug.

(Tweet 5): 3 cops shot dead in Baton Rouge. Shooter is black. Another #BlackLivesMatter-inspired attack, no doubt.

(Tweet 6): What is this world coming to when you can't aim a gun at some cops without them shooting you? #BlackLivesMatter

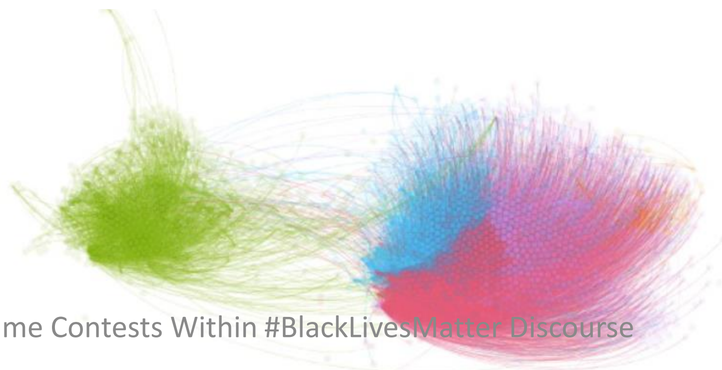
Stewart et al. (2017). Drawing the Lines of Contention: Networked Frame Contests Within #BlackLivesMatter Discourse

Networked Frame Contests within #BlackLivesMatter Discourse

- Study **framing contests** between left- and right-leaning super-clusters
- Defending and revising frames against challenges (left)

(Tweet 7 - Left Leaning): Question of the day to #BlueLivesMatter~ Does the police shooting of #CharlesKinsey hurt your cause?

(Tweet 8 - Left Leaning): WHERE'S ALL THE #BlueLivesMatter PEOPLE?? 2 POLICE OFFICERS SHOT BY 2 WHITE MEN, BOTH SHOOTERS IN CUSTODY NOT DEAD.



Stewart et al. (2017). Drawing the Lines of Contention: Networked Frame Contests Within #BlackLivesMatter Discourse

Networked Frame Contests within #BlackLivesMatter Discourse

- Study **framing contests** between left- and right-leaning super-clusters
- Defending and revising frames against challenges (right)

```
(Tweet 9 - Right Leaning): A 2-year-old girl was shot in the
head Friday in a drive-by shooting in Cleveland -
#BlackLivesMatter DO YOU CARE???
```

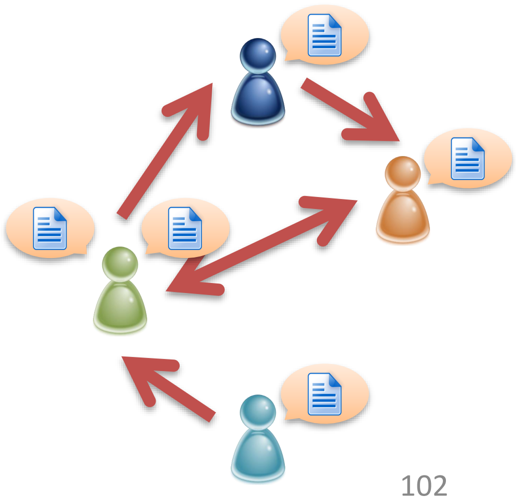
```
(Tweet 11 - Right Leaning): How is shooting cops in Dallas
justice for whatever may have happened elsewhere? It is not.
#BlueLivesMatter
```

Stewart et al. (2017). Drawing the Lines of Contention: Networked Frame Contests Within #BlackLivesMatter Discourse

Online Debate Forums

- Social media sites for debating issues

- Valuable resources for:
  - Argumentation
  - Dialogue
  - Sentiment
  - Opinion mining



CreateDebate.org

Barack Obama VS Mitt Romney

Barack Obama — Side Score: 15 — ADD ARGUMENT

VS

Mitt Romney — Side Score: 16 — ADD ARGUMENT

Cuaroc (5871) — 1 point

Well Obama won didn't he?

2 years ago Side: Barack Obama

Support Dispute Clarify

↑ Hide Replies

> **GuitaristDog (2543) Disputed** — 2 points
>
> Woodrow Wilson won, he passed an amendment allowing for the government to tax by "any means necessary" to get around the constitutions protection on individual liberties, passed the Federal Reserve Act and created the IRS. He also imprisoned many people who were against/protested him, including war protesters and women's suffragists.
>
> I wouldn't consider where someone won an election or not to be the judge of how good they are.
>
> 2 years ago Side: Mitt Romney
>
> Support Dispute Clarify

> **Cuaroc (5871)** — 2 points
>
> looks like people don't like the truth.
>
> 2 years ago Side: Barack Obama
>
> Support Dispute Clarify

---

debateking1 (13) — 3 points

Mitt would certainly have risen up USA's economy as he did to Massachusetts

2 years ago Side: Mitt Romney

Support Dispute Clarify

↑ Hide Replies

> **ASHenry (6) Disputed** — 2 points
>
> Mass. is 49th in state for job increase. That means out of 50 states, 50 being the worse you can do, Mass was number 49 under Romney.
>
> 2 years ago Side: Barack Obama
>
> Support Dispute Clarify

Theo (2) — 2 points

Barack winning is not a benefit for America... isn't it already pretty much a fact that he lied to the American people... what is he doing about Iran about to get nuclear weapons.... what is he doing to back Israel??? what is he doing for the economy other than over hiring??

2 years ago Side: Mitt Romney

Support Dispute Clarify

MrRight (67) — 2 points

CreateDebate.org

Debate topic

Barack Obama VS Mitt Romney

Barack Obama
Side Score: 15
ADD ARGUMENT

VS

Mitt Romney
Side Score: 16
ADD ARGUMENT

Cuaroc (5871) **1 point**

Well Obama won didn't he?

2 years ago Side: Barack Obama

Support Dispute Clarify

↑ Hide Replies

GuitaristDog (2543) Disputed **2 points**

Woodrow Wilson won, he passed an amendment allowing for the government to tax by "any means necessary" to get around the constitutions protection on individual liberties, passed the Federal Reserve Act and created the IRS. He also imprisoned many people who were against/protested him, including war protesters and women's suffragists.

I wouldn't consider where someone won an election or not to be the judge of how good they are.

2 years ago Side: Mitt Romney

Support Dispute Clarify

Cuaroc (5871) **2 points**

looks like people don't like the truth.

2 years ago Side: Barack Obama

Support Dispute Clarify

debateking1 (13) **3 points**

Mitt would certainly have risen up USA's economy as he did to Massachusetts

2 years ago Side: Mitt Romney

Support Dispute Clarify

↑ Hide Replies

ASHenry (6) Disputed **2 points**

Mass. is 49th in state for job increase. That means out of 50 states, 50 being the worse you can do, Mass was number 49 under Romney.

2 years ago Side: Barack Obama

Support Dispute Clarify

Theo (2) **2 points**

Barack winning is not a benefit for America... isn't it already pretty much a fact that he lied to the American people... what is he doing about Iran about to get nuclear weapons.... what is he doing to back Israel??? what is he doing for the economy other than over hiring??

2 years ago Side: Mitt Romney

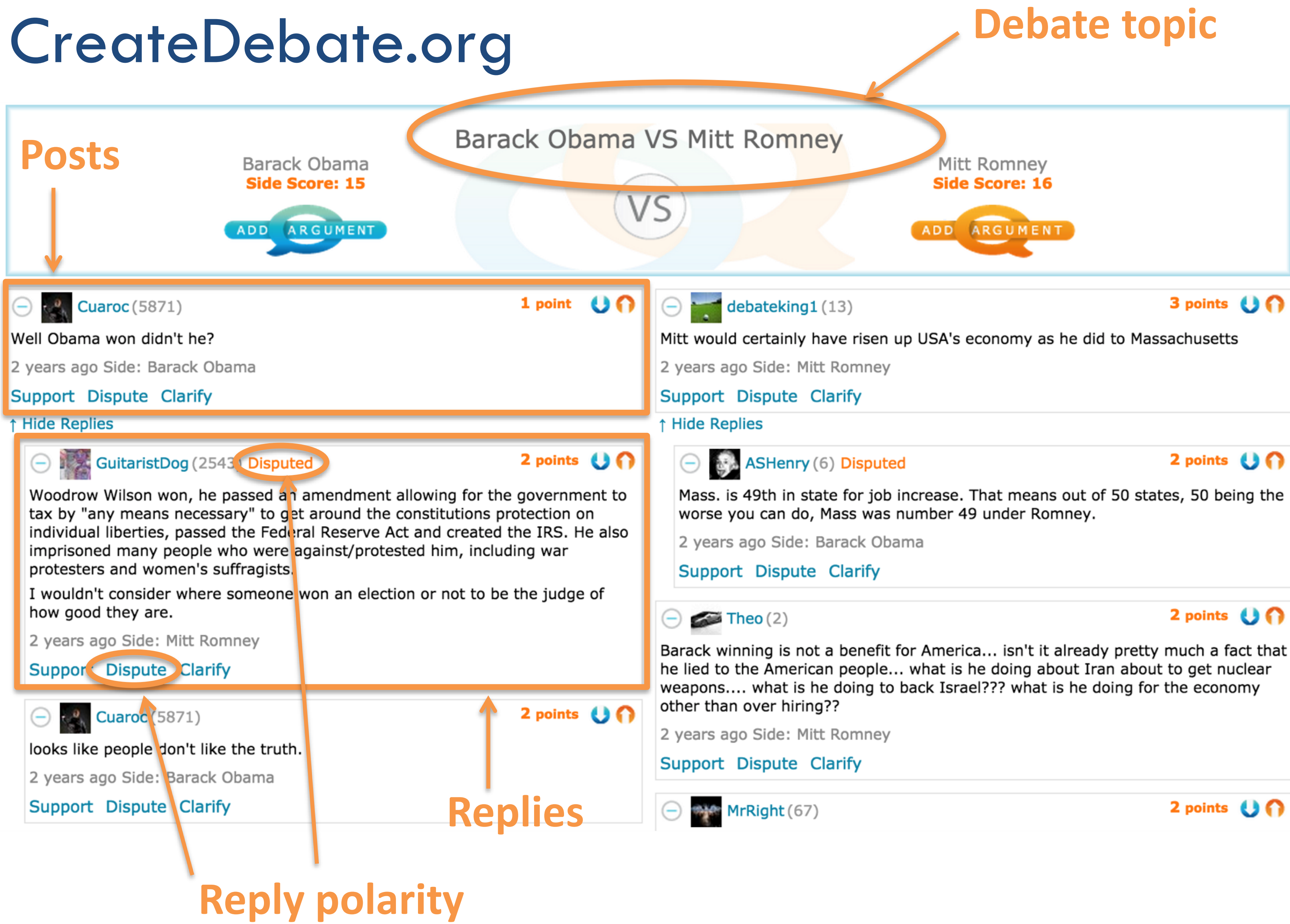
Support Dispute Clarify

MrRight (67) **2 points**

CreateDebate.org

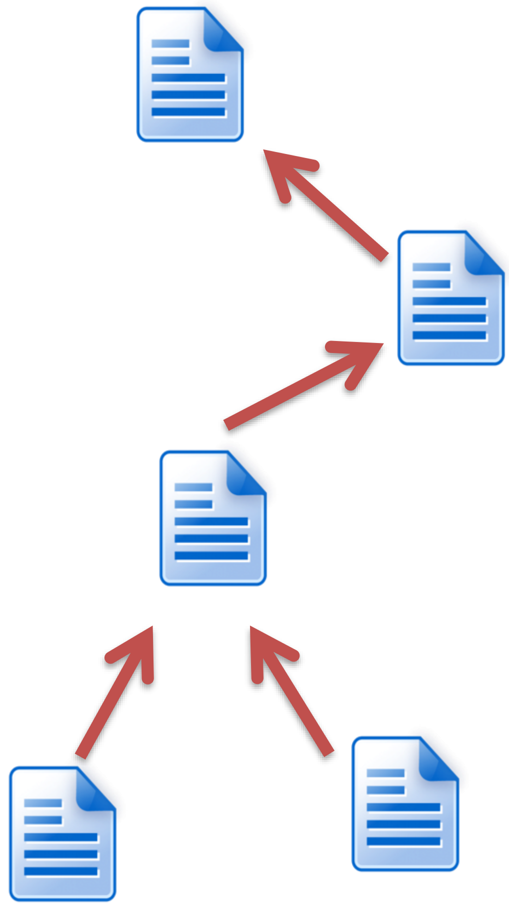
CreateDebate.org

CreateDebate.org

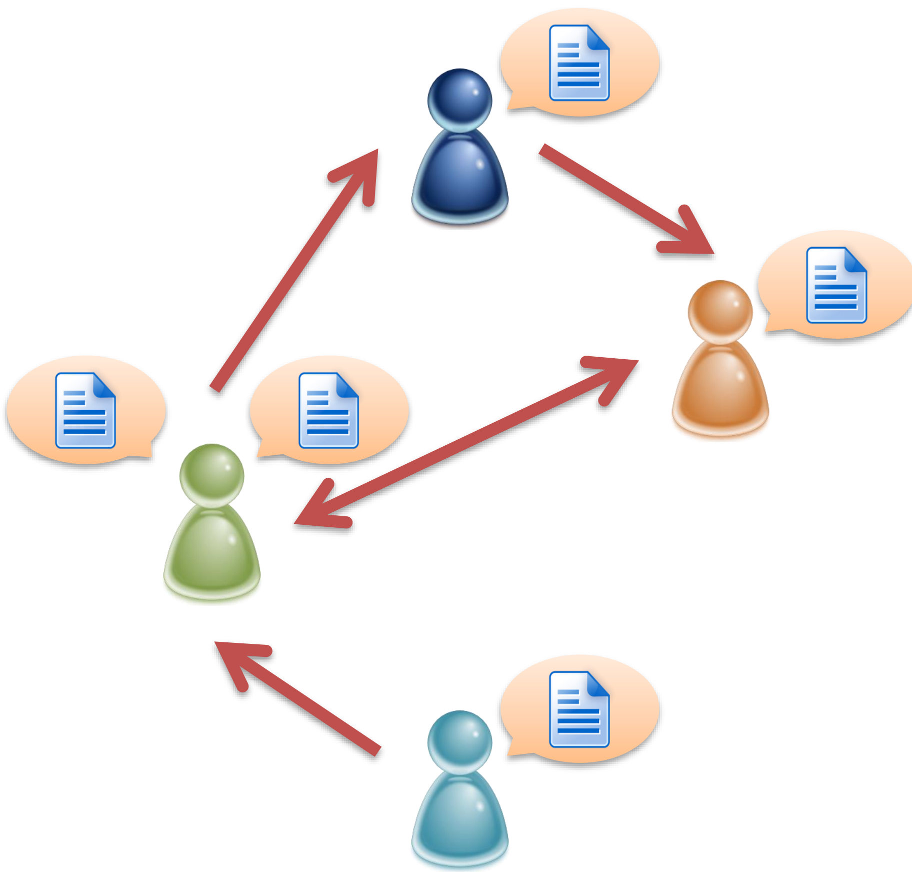


Online Debate Forums

Graph of posts:
tree structure

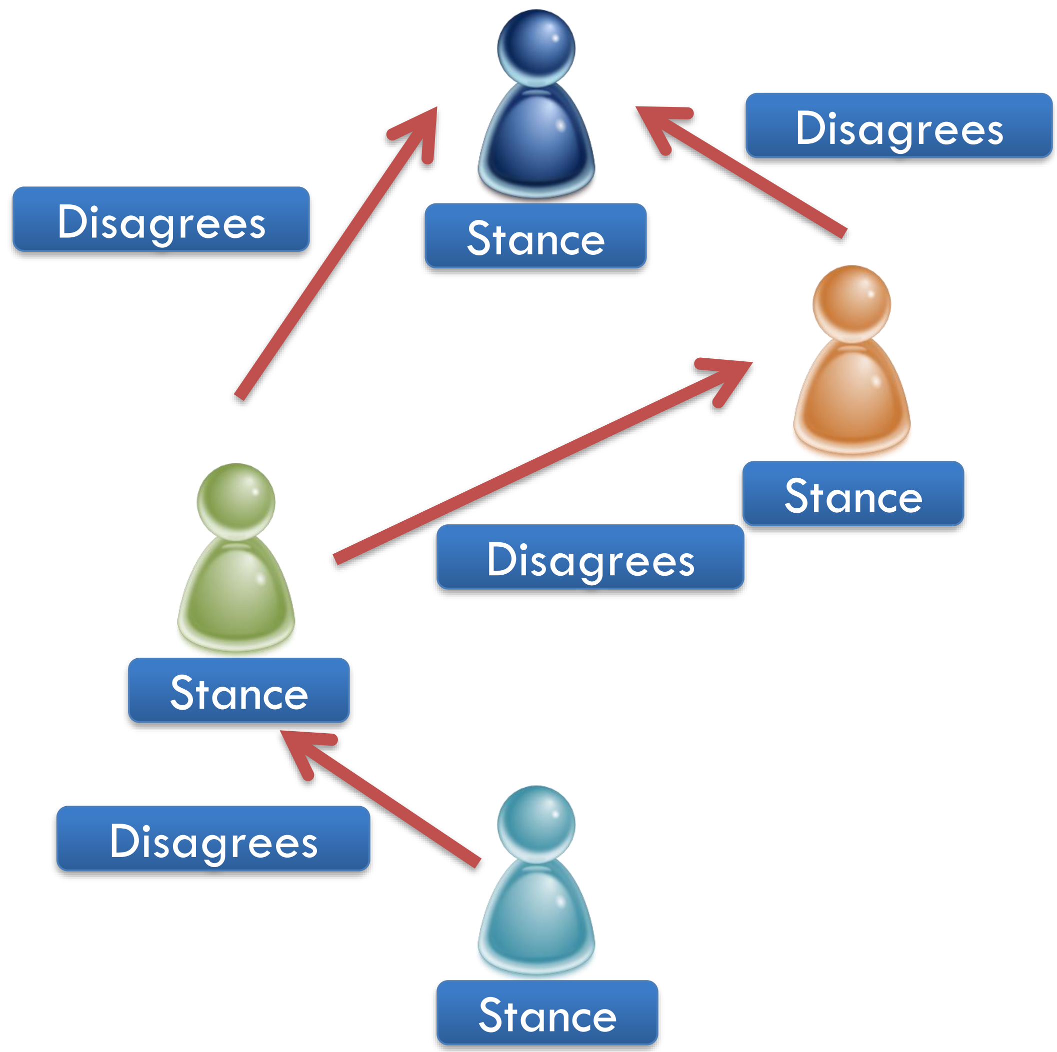


Graph of users:
loopy structure



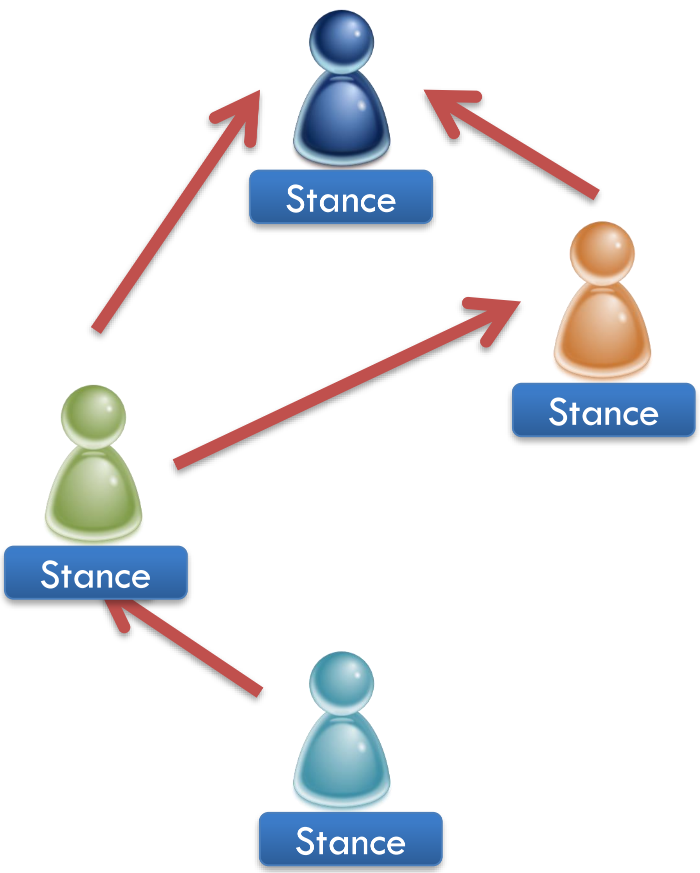
Classification Targets

- **Stance**
  - Author-level
  - Post-level
- **Disagreement**
  - Author-level
  - Post-level
  - Textual

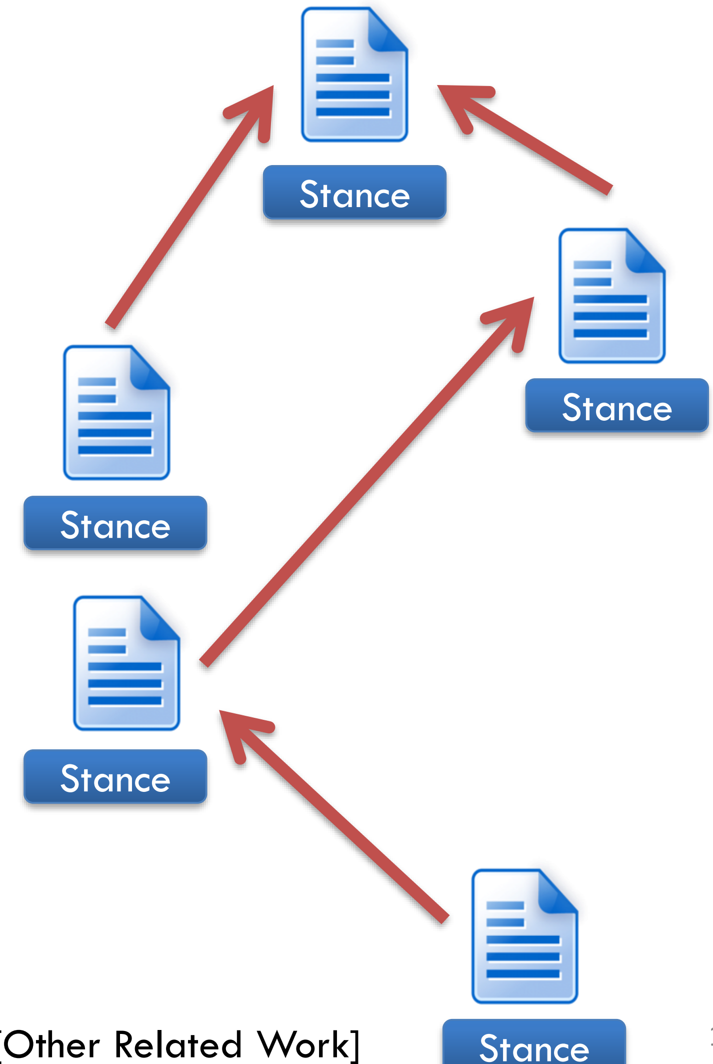


Modeling Question 1)

Modeling at author-level or post-level?



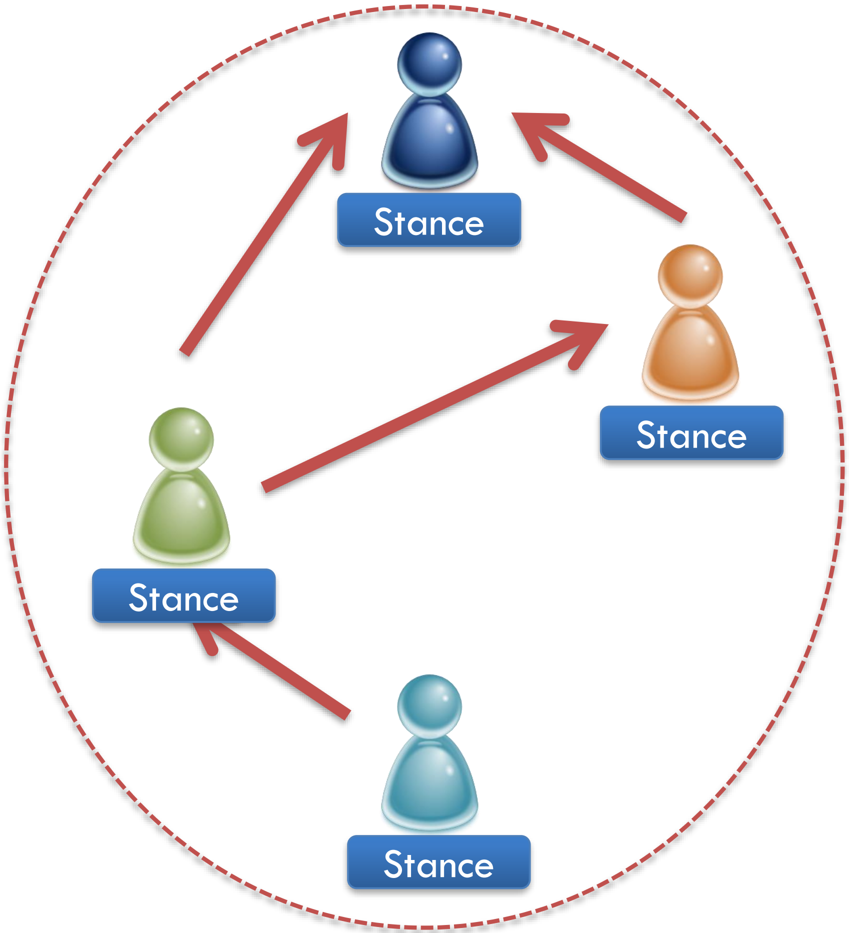
[Hasan and Ng 2013]



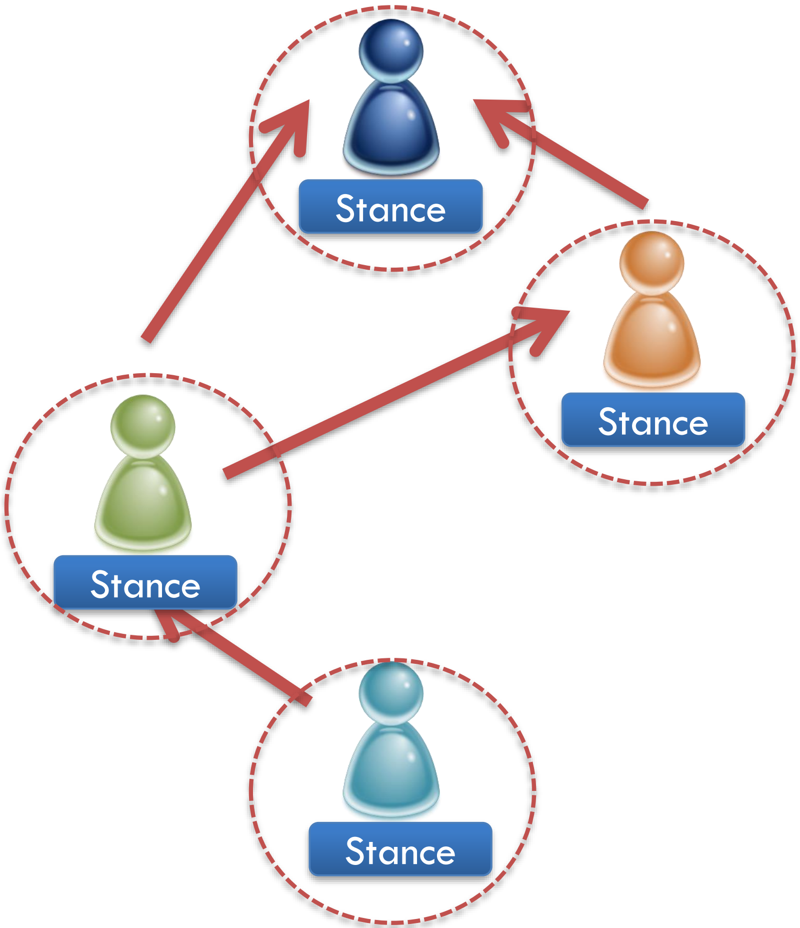
[Other Related Work]

Modeling Question 2)

Collective classification vs. local classification?



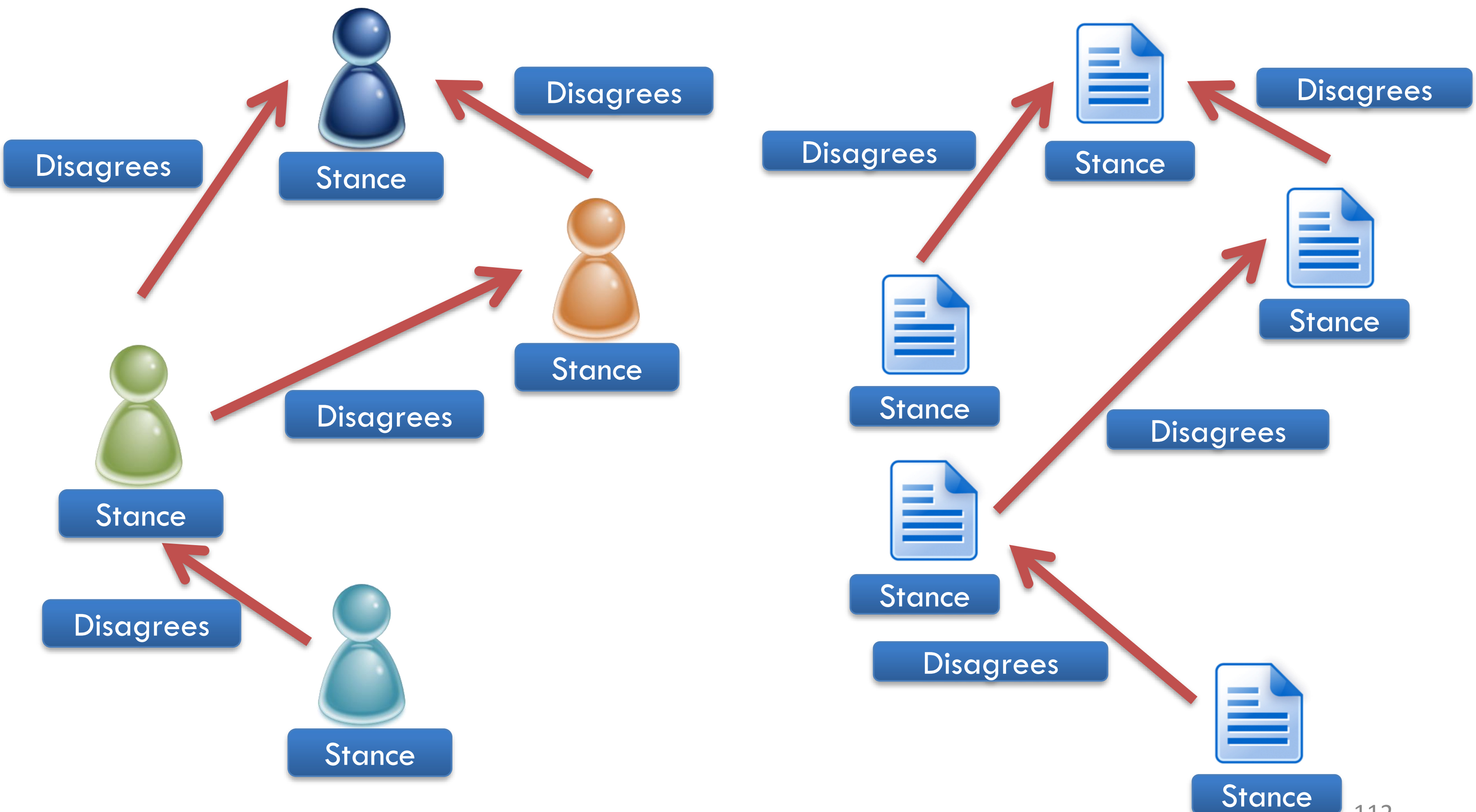
[Walker et al. 2012, Hasan and Ng 2013]



[Walker et al. 2012]

Modeling Question 3)

Jointly model disagreement together with stance?



[Abbott et. al 2012 - Linguistic Features], [Burfoot et. al 2011 for Congressional Debates]

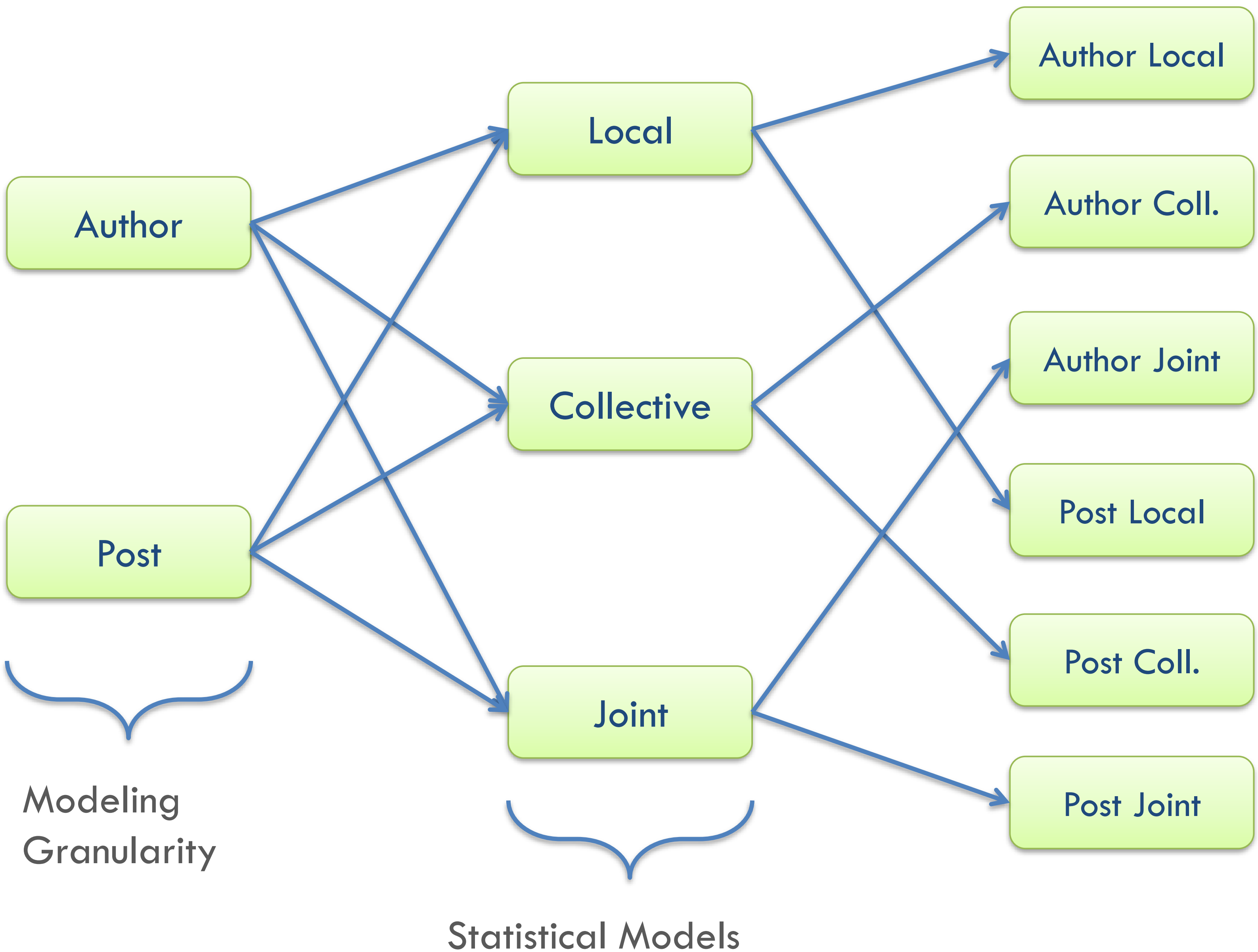
Our Contributions

- A **unified framework** to explore multiple models

- Fast, highly **scalable inference**
  - Large post-level graphs
  - Loopy author-level graphs

- **Systematic study** of modeling options
  - Modeling recommendations

All Combinations of Models



Probabilistic Soft Logic (PSL)

- Templating language for highly scalable graphical model called Hinge-loss Markov Random Fields

Relaxations of Logical Operators

$$5.0: Disagrees(A1, A2) \wedge Pro(A1) \rightarrow \sim Pro(A2)$$

Rule Weight

Predicates are continuous Random Variables!

Hinge-loss MRFs Over Continuous Variables

$$\underbrace{P(\mathbf{Y}|\mathbf{X})}_{\text{Conditional random field over continuous RVs in [0,1]}} \propto \exp\underbrace{\left(-\sum_{j=1}^{M}\lambda_j\psi_j(\mathbf{X},\mathbf{Y})\right)}_{\text{Feature function for each instantiated rule}}$$

5.0: Disagrees( *,*  *) ^ Pro(*  *) → ~Pro(*  *)*

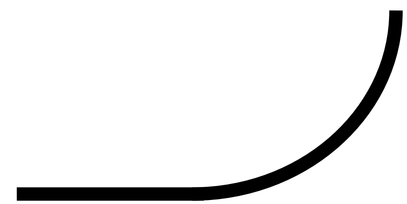
Hinge-loss MRFs Over Continuous Variables

$$P(\mathbf{Y}|\mathbf{X}) \propto \exp\left(-\sum_{j=1}^{M} \lambda_j \underline{\psi_j(\mathbf{X}, \mathbf{Y})}\right)$$

Feature functions are hinge loss functions

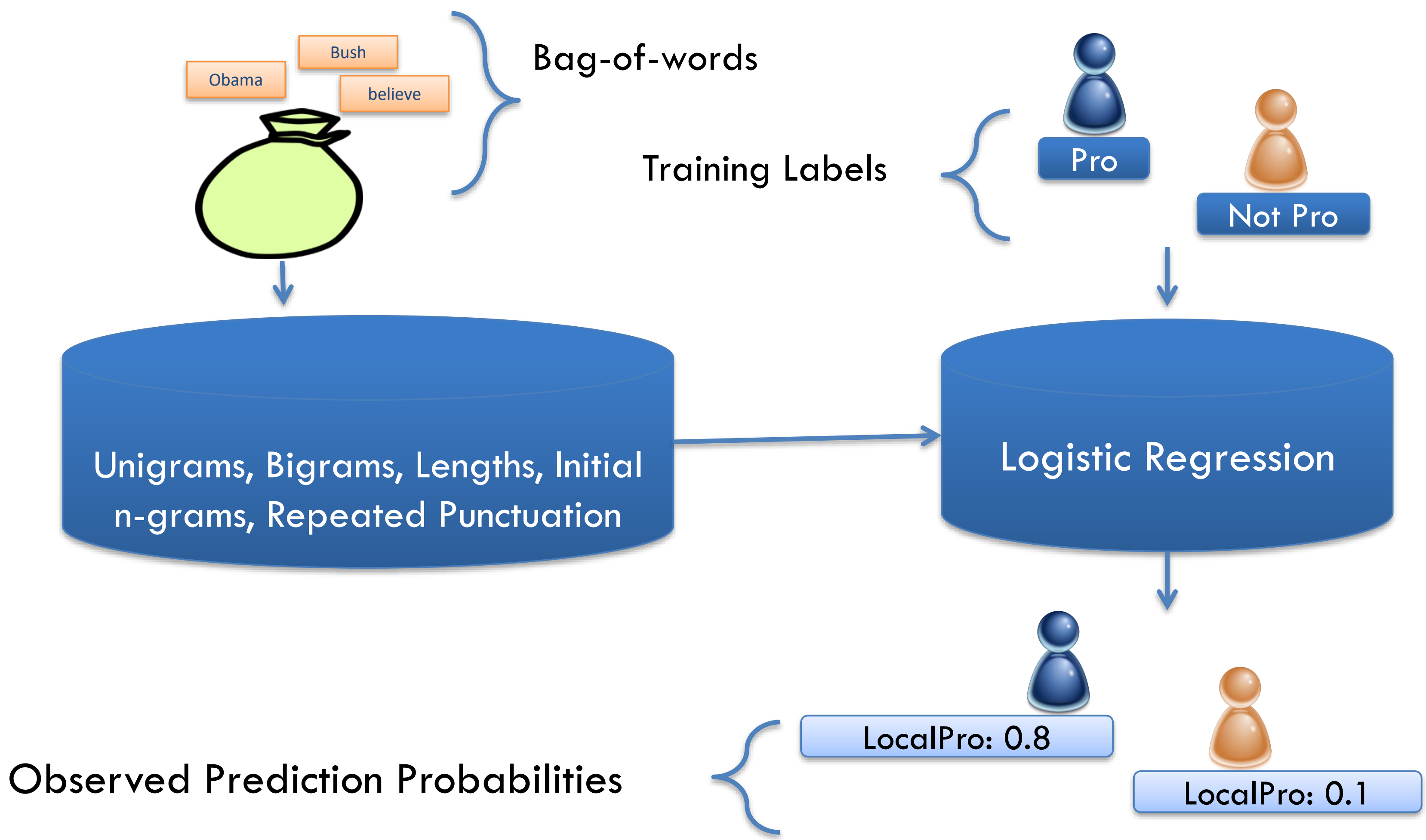
$$\psi_j(\mathbf{X}, \mathbf{Y}) = \max\{\underline{l_j(\mathbf{X}, \mathbf{Y})}, 0\}^2$$

Linear function



Hinge losses encode the *distance to satisfaction* for each instantiated rule

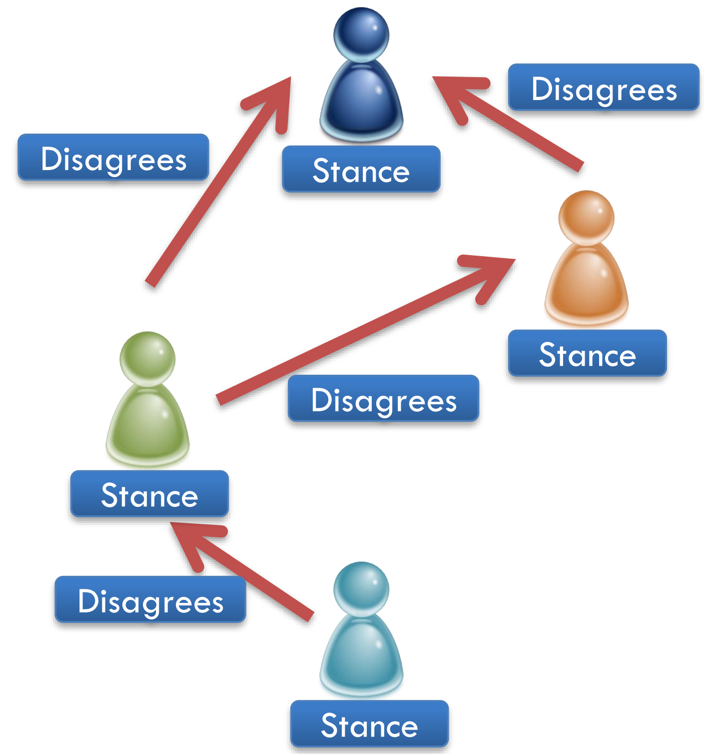
Constructing Local Predictors



PSL Rules for Stance Prediction Models

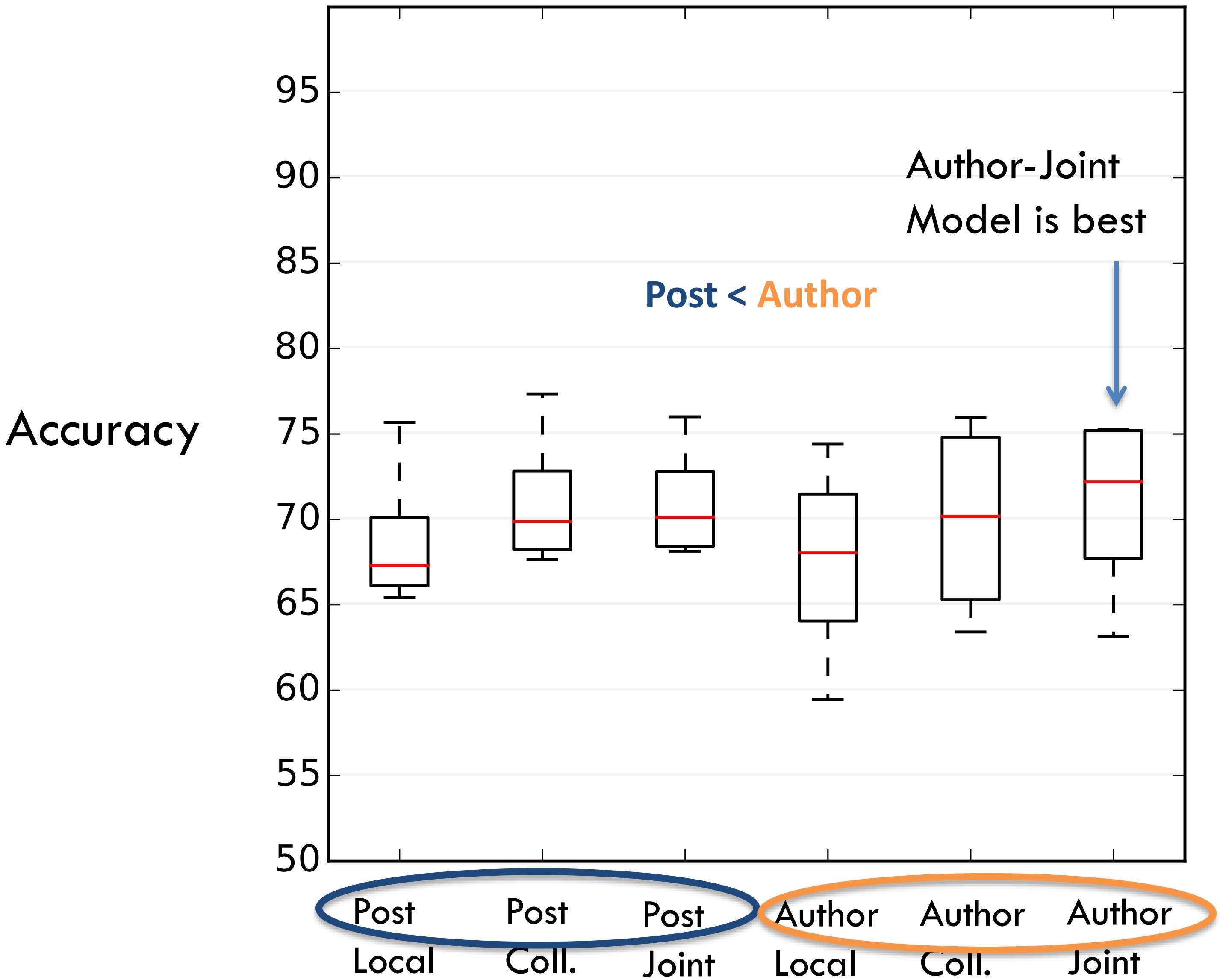
- Local classifiers for **stance** (e.g. pro gun control)
- Local classifiers for **disagreement**
- **Collective classification** on stance and disagreement
  - Can model either at **author** or **post** level

- Three increasingly complicated models:
  - Just **local prediction**
  - **Collective**, reply edge implies reverse polarity
  - **Disagreement** modeling

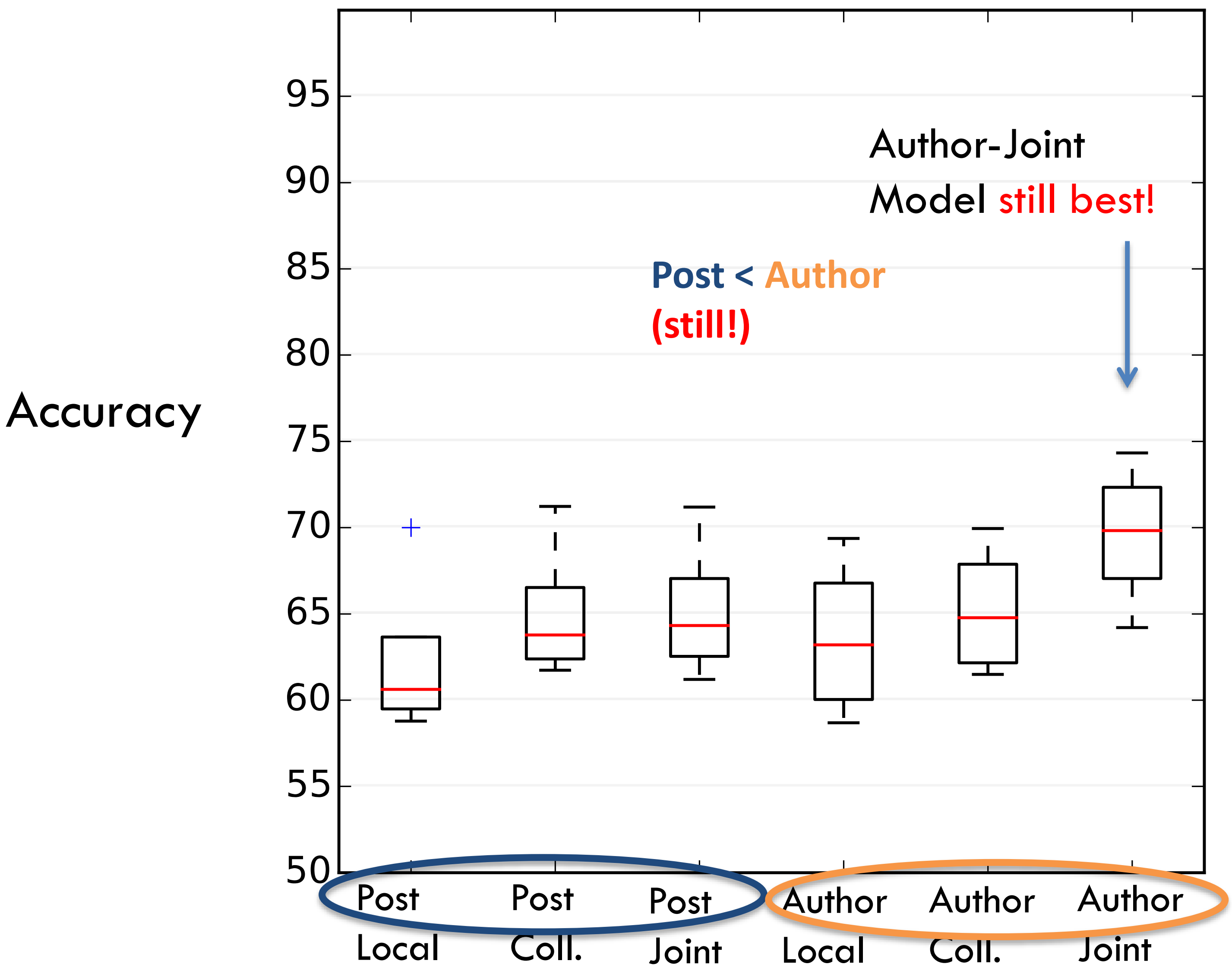


All models:		Collective models only:		Disagreement models only:	
$\mathit{localPro}(\mathrm{X1})$	$\rightarrow \mathit{pro}(\mathrm{X1})$	$\mathit{disagree}(\mathrm{X1}, \mathrm{X2}) \land \mathit{pro}(\mathrm{X1})$	$\rightarrow \neg\, \mathit{pro}(\mathrm{X2})$	$\mathit{localDisagree}(\mathrm{X1}, \mathrm{X2})$	$\rightarrow \mathit{disagree}(\mathrm{X1}, \mathrm{X2})$
$\neg\, \mathit{localPro}(\mathrm{X1})$	$\rightarrow \neg\, \mathit{pro}(\mathrm{X1})$	$\mathit{disagree}(\mathrm{X1}, \mathrm{X2}) \land \neg\, \mathit{pro}(\mathrm{X1})$	$\rightarrow \mathit{pro}(\mathrm{X2})$	$\neg\, \mathit{localDisagree}(\mathrm{X1}, \mathrm{X2})$	$\rightarrow \neg\, \mathit{disagree}(\mathrm{X1}, \mathrm{X2})$
		$\neg\, \mathit{disagree}(\mathrm{X1}, \mathrm{X2}) \land \mathit{pro}(\mathrm{X1})$	$\rightarrow \mathit{pro}(\mathrm{X2})$	$\mathit{pro}(\mathrm{X1}) \land \neg\, \mathit{pro}(\mathrm{X2})$	$\rightarrow \mathit{disagree}(\mathrm{X1}, \mathrm{X2})$
		$\neg\, \mathit{disagree}(\mathrm{X1}, \mathrm{X2}) \land \neg\, \mathit{pro}(\mathrm{X1})$	$\rightarrow \neg\, \mathit{pro}(\mathrm{X2})$	$\mathit{pro}(\mathrm{X1}) \land \mathit{pro}(\mathrm{X2})$	$\rightarrow \neg\, \mathit{disagree}(\mathrm{X1}, \mathrm{X2})$
		$\mathit{disagree}(\mathrm{X1}, \mathrm{X2})$	$= 1$	$\neg\, \mathit{pro}(\mathrm{X1}) \land \neg\, \mathit{pro}(\mathrm{X2})$	$\rightarrow \neg\, \mathit{disagree}(\mathrm{X1}, \mathrm{X2})$

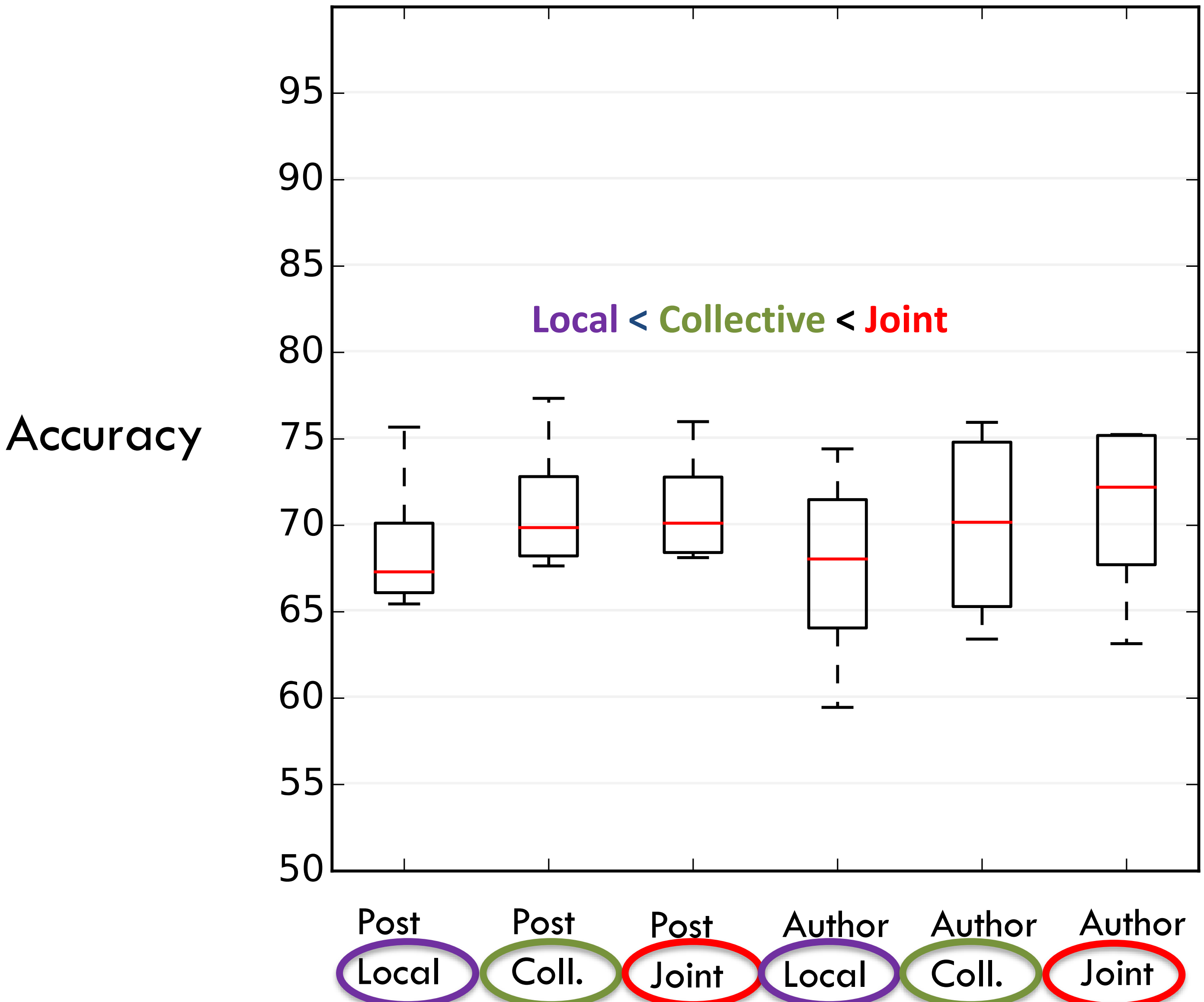
Author Stance Prediction – CreateDebate.org



Post Stance Prediction – CreateDebate.org



Author Stance Prediction – CreateDebate.org



Modeling Influence Relationships in the U.S. Supreme Court

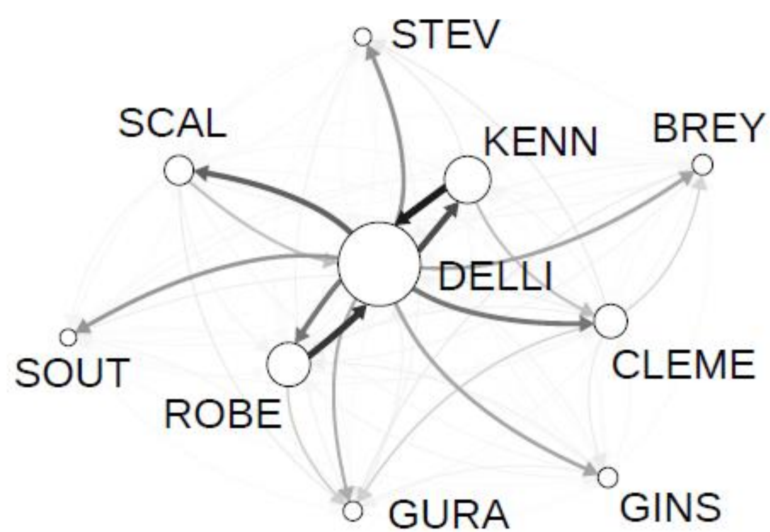
The Bayesian Echo Chamber: Modeling Social Influence via Linguistic Accommodation

Fangjian Guo
Duke University
Durham, NC, USA
guo@cs.duke.edu

Charles Blundell
Gatsby Unit, UCL
London, UK
c.blundell@gatsby.ucl.ac.uk

Hanna Wallach
Microsoft Research
New York, NY, USA
wallach@microsoft.com

Katherine Heller
Duke University
Durham, NC, USA
kheller@stat.duke.edu



Guo, F., Blundell, C., Wallach, H., and Heller, K. (2015). AISTATS

Modeling Influence Relationships in the U.S. Supreme Court

- Model intuition: **linguistic accommodation**
  - Influential speakers lead others to **use the same words as them**
  - Weighted **influence network** $\rho^{(qp)}$ determines influence relationships

- Infer influence network via **Bayesian inference**

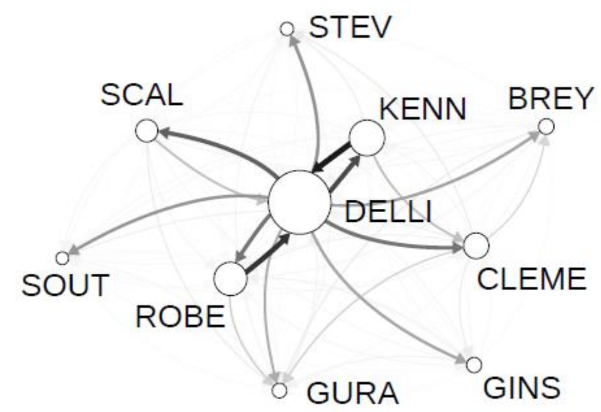
Expected word probabilities, person p, word v, utterance n

Word counts for person q, with time decay

$$B_{v,n}^{(p)} \propto \beta_v^{(p)} + \sum_{q \neq p} \rho^{(qp)} \psi_{v,n}^{(qp)}$$

Person p's inherent language usage

Influence from person q to person p

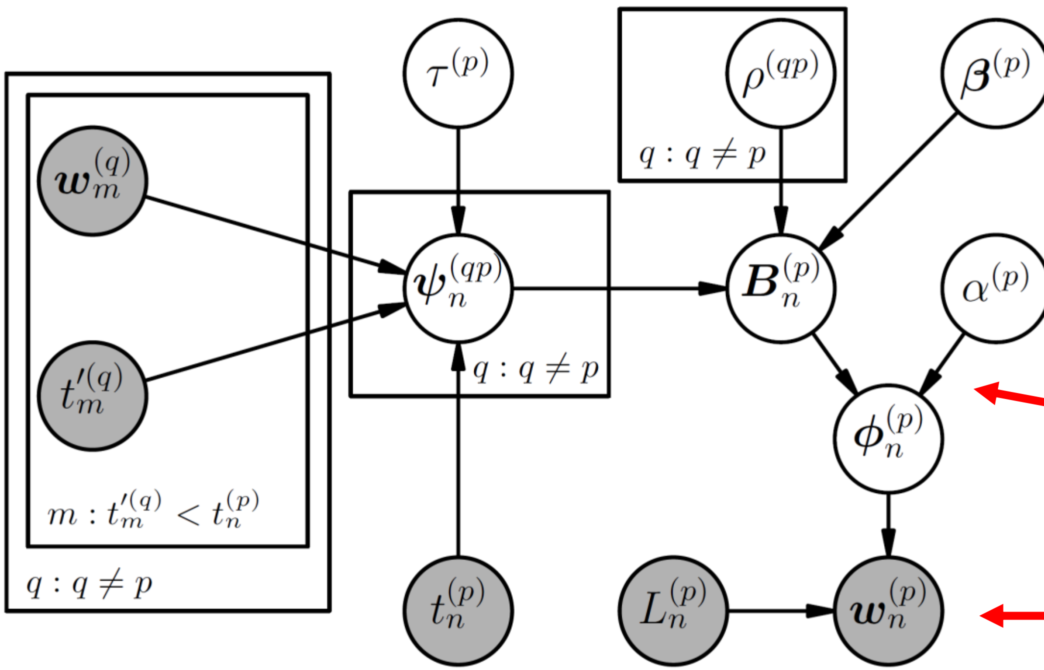


Guo, F., Blundell, C., Wallach, H., and Heller, K. (2015). AISTATS

Modeling Influence Relationships in the U.S. Supreme Court

Influence from person q to person p

Previous utterances and their end times



$$B_{v,n}^{(p)} \propto \beta_v^{(p)} + \sum_{q \neq p} \rho^{(qp)} \psi_{v,n}^{(qp)}$$

Dirichlet distribution (allows the final word distribution $\boldsymbol{\phi}^{(p)}$ to deviate from $\boldsymbol{B}^{(p)}$)

Person p's nth utterance: timestamp t, length L, words w

Word counts (time decayed)

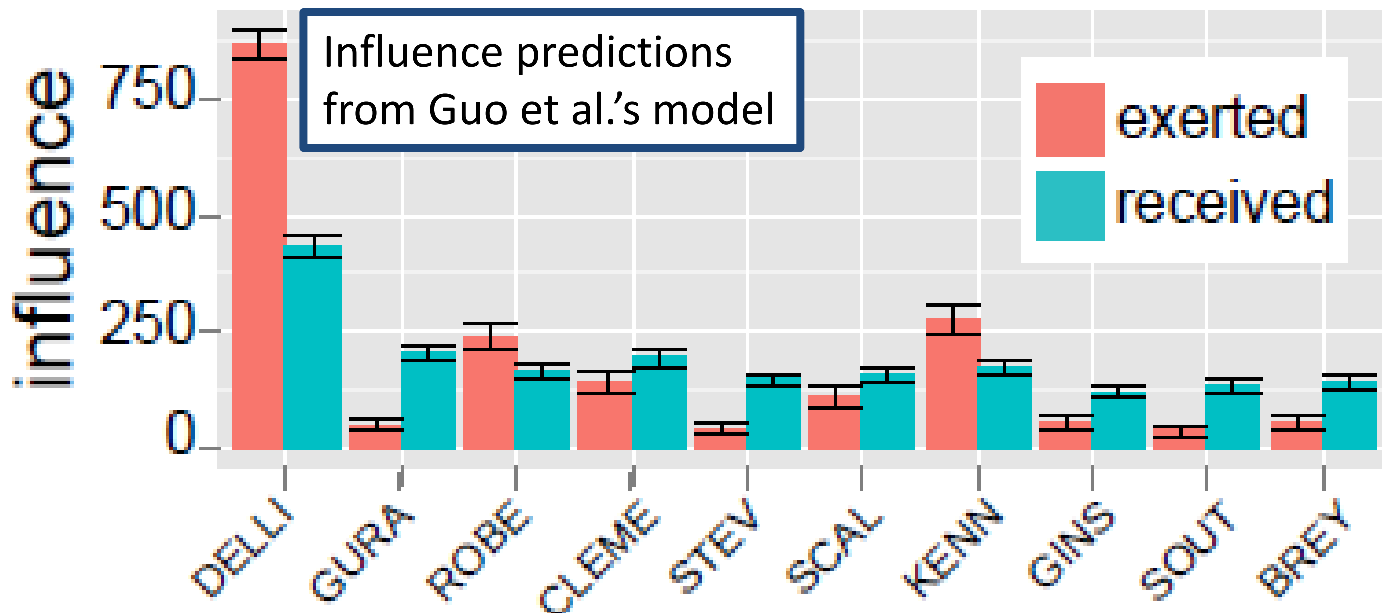
$$\psi_{v,n}^{(qp)} = \sum_{m : t'^{(q)}_m < t_n^{(p)}} \left(\sum_{l=1}^{L_m^{(q)}} \mathbf{1}(w_{l,m}^{(q)} = v) \right) \exp\left(- \frac{t_n^{(p)} - t'^{(q)}_m}{\tau_L^{(p)}} \right)$$

Time decay

Guo, F., Blundell, C., Wallach, H., and Heller, K. (2015). AISTATS

Modeling Influence Relationships in the U.S. Supreme Court

Total influence exerted and received, District of Columbia v. Heller case



Represented petitioner

Represented respondent

Supreme court justices

Guo, F., Blundell, C., Wallach, H., and Heller, K. (2015). AISTATS

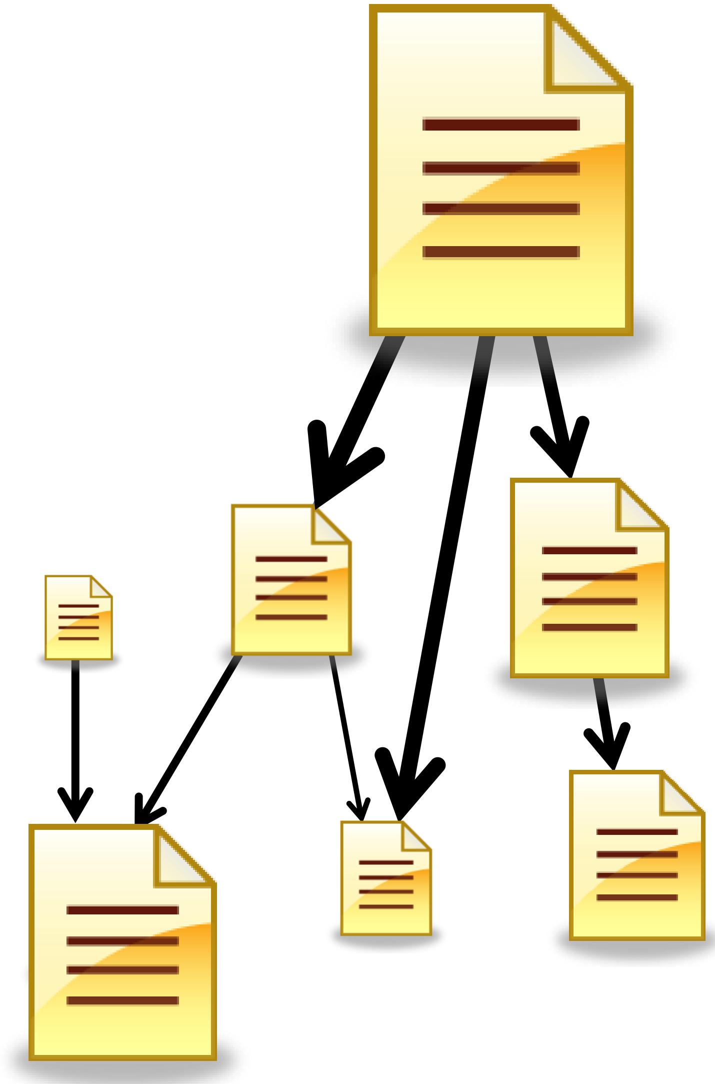
Modeling Influence in Citation Networks

Which are the most **important articles?**

What are the **influence relationships** between articles?

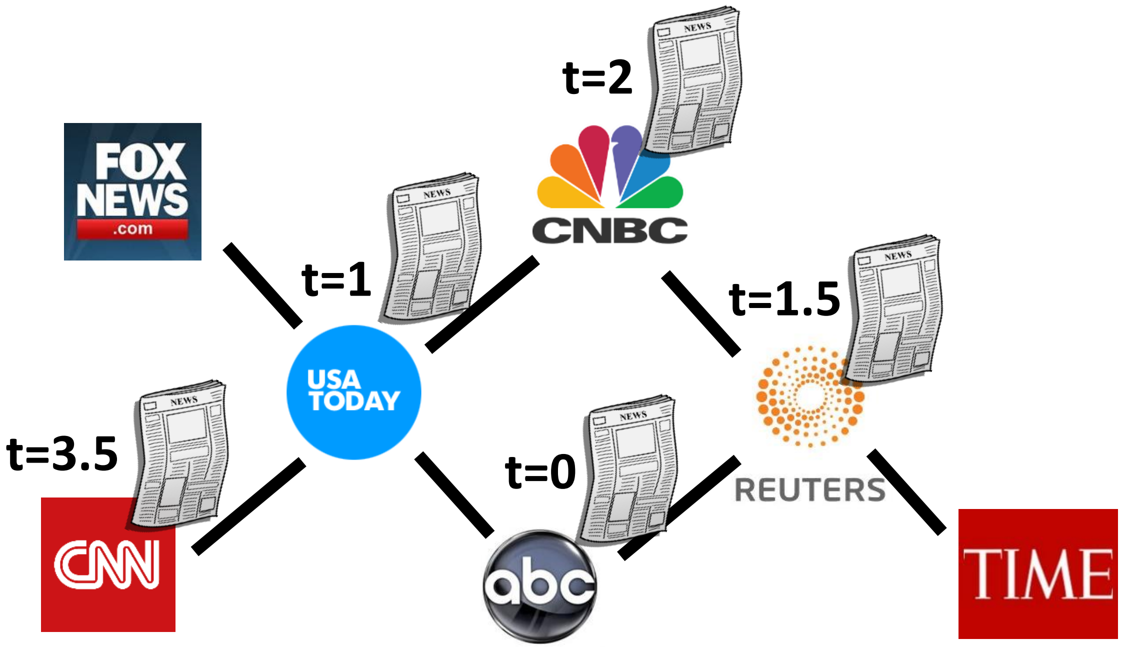
A similar model can be used in this context as well (Foulds and Smyth, 2013)

- Dirichlet priors cause **influenced documents** to **accommodate topics** instead of words



Foulds and Smyth (2013). Modeling Scientific Impact with Topical Influence Regression. EMNLP

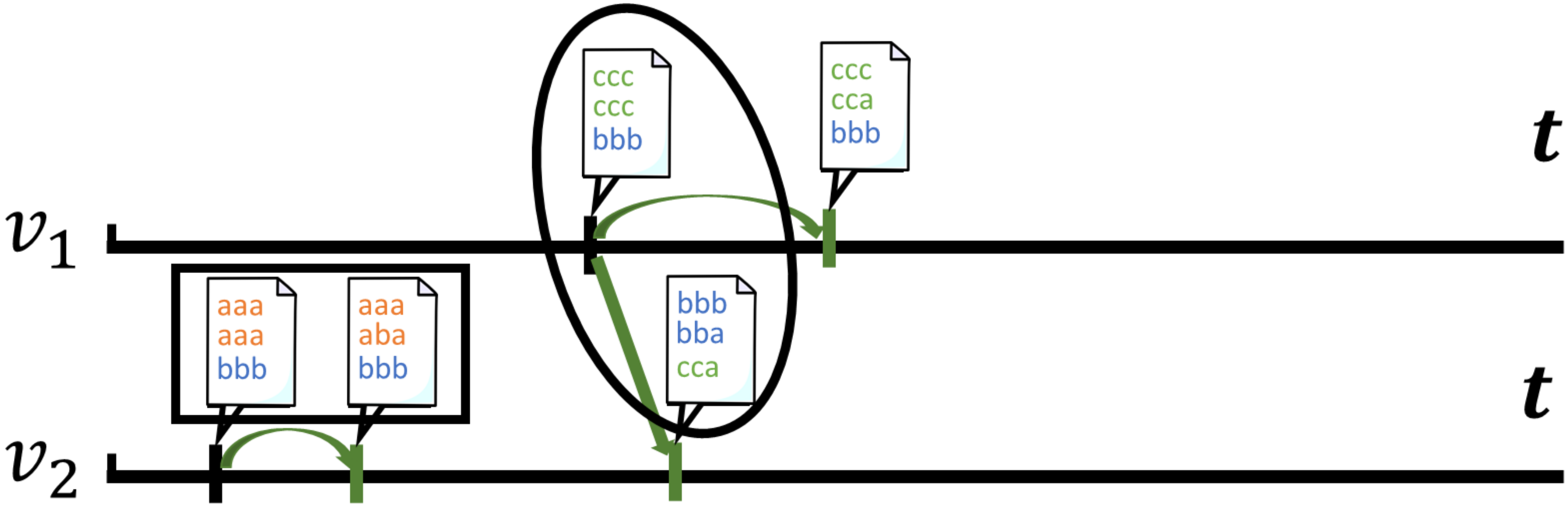
Information diffusion in text-based cascades



- Temporal information
- Content information
- Network is latent

X. He, T. Rekatsinas, J. R. Foulds, L. Getoor, and Y. Liu. HawkesTopic: A joint model for network inference and topic modeling from text-based cascades. ICML 2015.

HawkesTopic model for text-based cascades



Mutual exciting nature: A posting event can trigger future events

Content cascades: The content of a document should be similar to the document that triggers its publication

X. He, T. Rekatsinas, J. R. Foulds, L. Getoor, and Y. Liu. HawkesTopic: A joint model for network inference and topic modeling from text-based cascades. ICML 2015.

Modeling posting times

Mutually exciting nature captured via
Multivariate Hawkes Process (MHP) [Liniger 09].

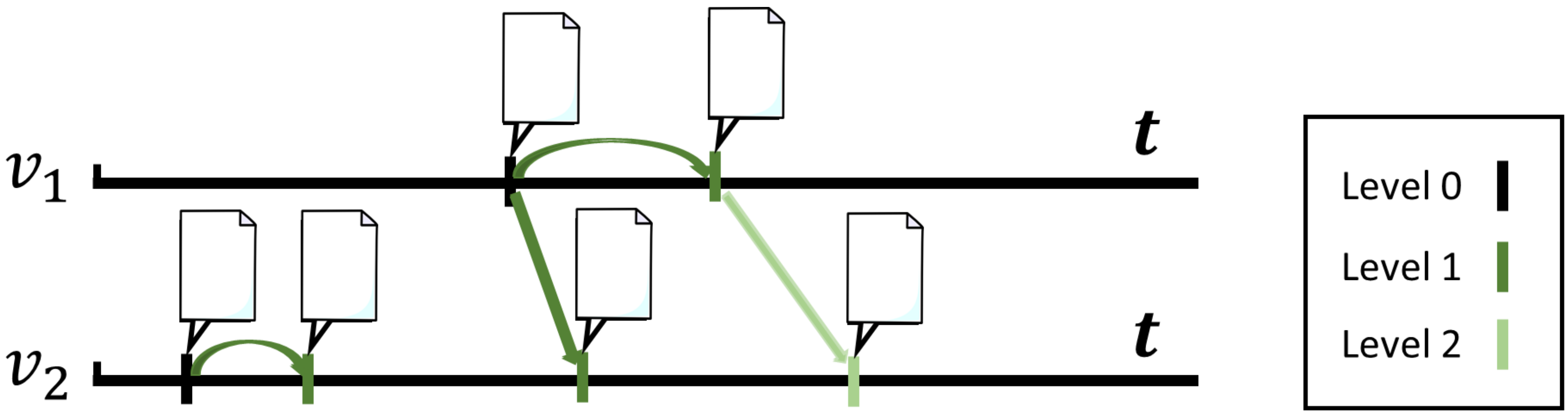
For MHP, **intensity process** $\lambda_v(t)$ takes the form:

Rate = Base intensity + Influence from previous events

$$\lambda_v(t) = \overbrace{\mu_v} + \overbrace{\textstyle\sum_{e:t_e<t} A_{v_e,v} f_\Delta(t - t_e)}$$

$A_{u,w}$: **influence strength** from u to v
$f_\Delta(\cdot)$: probability density function of the **delay distribution**

Clustered Poisson process interpretation

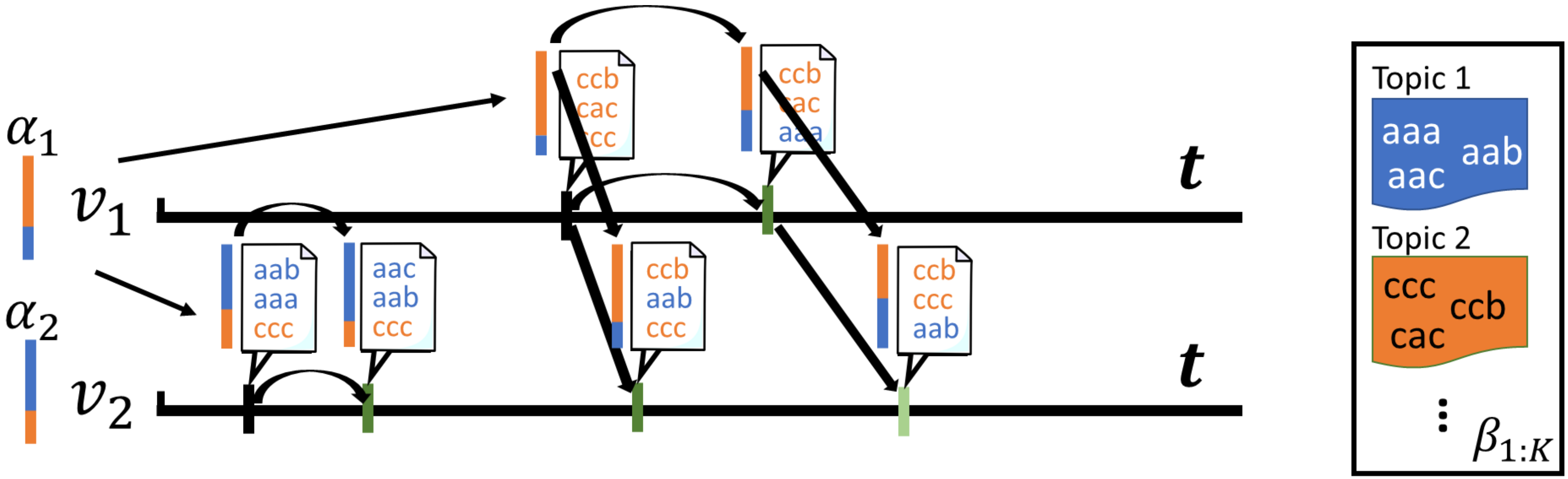


Generate events and their posting times in a **breadth first** order by interpreting the MHP as **clustered Poisson process** [Simma 10]

Provide explicit **parent relationship** for evolution of the content information

X. He, T. Rekatsinas, J. R. Foulds, L. Getoor, and Y. Liu. HawkesTopic: A joint model for network inference and topic modeling from text-based cascades. ICML 2015.

Generating documents



Step 1: Generate the topics $\beta_{1:K}$: $\beta_k \sim Dir(\alpha)$

Step 2: For spontaneous events (level=0): $\eta_e \sim N(\alpha_v, \sigma^2 I)$

Step 3: For triggered events (level>0): $\eta_e \sim N(\eta_{\text{parent}[e]}, \sigma^2 I)$

Step 4: For each word in each document: $z_{e,n} \sim \text{Discrete}\big(\pi(\eta_e)\big), x_{e,n} \sim \text{Discrete}(\beta_{z_{e,n}})$

X. He, T. Rekatsinas, J. R. Foulds, L. Getoor, and Y. Liu. HawkesTopic: A joint model for network inference and topic modeling from text-based cascades. ICML 2015.

Experiments for HawkesTopic

"Ebola" news articles ~4 months
~9k articles, 330 news media sites
Copying information as ground truth

High-energy physics theory papers ~12 years
Top 50/100/200 researchers
Citation network as ground truth

Evaluation metrics:
-- Topic modeling: document competition likelihood [Wallach et al. 09]
-- Network Inference: AUC against the ground truth network

X. He, T. Rekatsinas, J. R. Foulds, L. Getoor, and Y. Liu. HawkesTopic: A joint model for network inference and topic modeling from text-based cascades. ICML 2015.

Results: ArXiv

Network Inference accuracy: 40% improvement

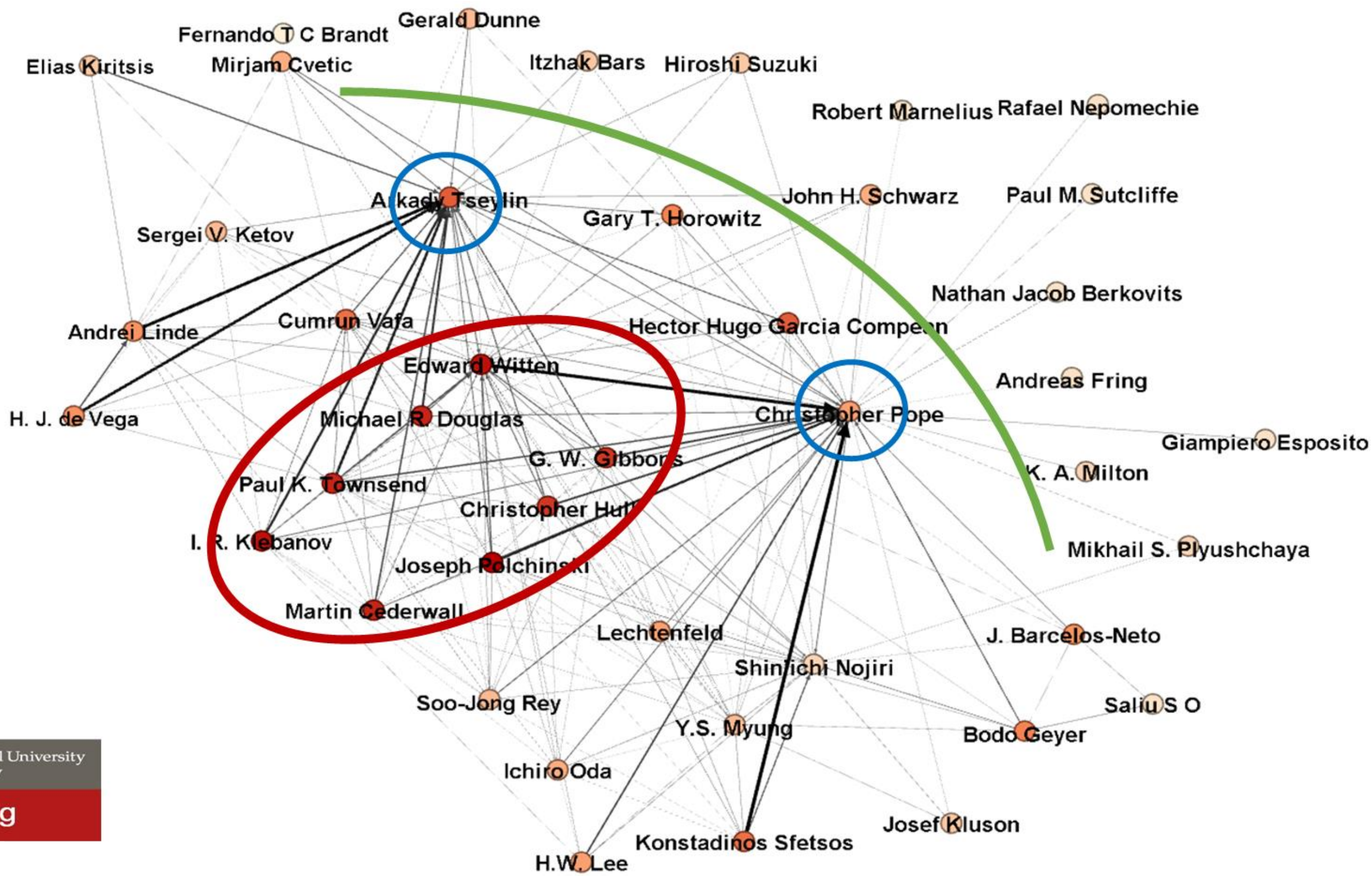
	Hawkes	Hawkes-LDA	Hawkes-CTM	HTM
Top50	0.594	0.656	0.645	**0.807**
Top100	0.588	0.589	0.614	**0.687**
Top200	0.618	0.630	0.629	**0.659**

Topic modeling accuracy:

	LDA	CTM	HTM
Top50	-11074	-10769	**-10708**
Top100	-15711	-15477	**-15252**
Top200	-27758	-27630	**-27443**

X. He, T. Rekatsinas, J. R. Foulds, L. Getoor, and Y. Liu. HawkesTopic: A joint model for network inference and topic modeling from text-based cascades. ICML 2015.

Results: ArXiv

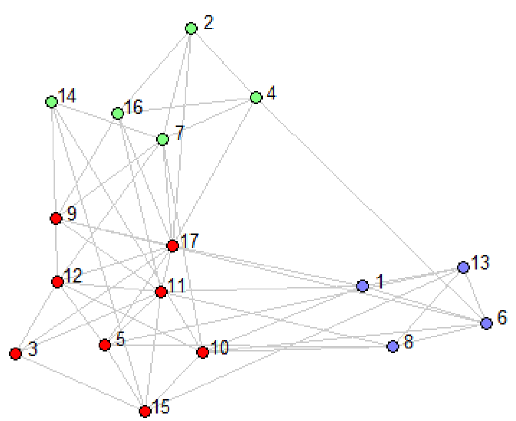


Cornell University Library
arXiv.org

X. He, T. Rekatsinas, J. R. Foulds, L. Getoor, and Y. Liu. HawkesTopic: A joint model for network inference and topic modeling from text-based cascades. ICML 2015.

Dynamic social network

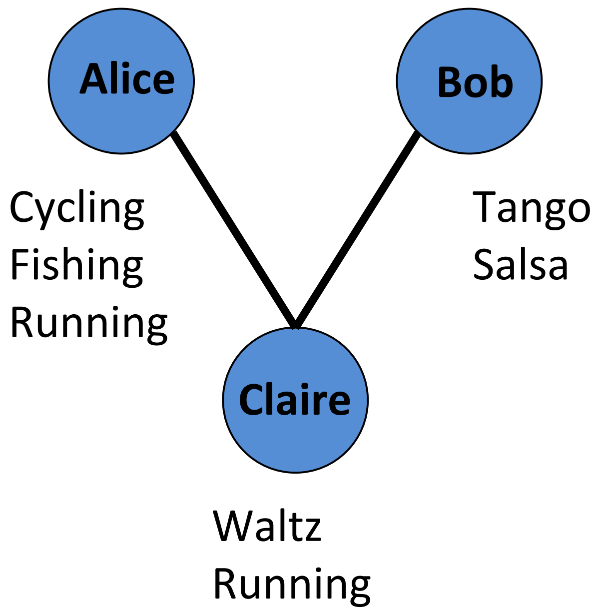
- Relations between people may change over time
- Need to generalize social network models to account for dynamics



Dynamic social network
(Nordlie, 1958; Newcomb, 1961)

Dynamic Relational Infinite Feature Model (DRIFT)

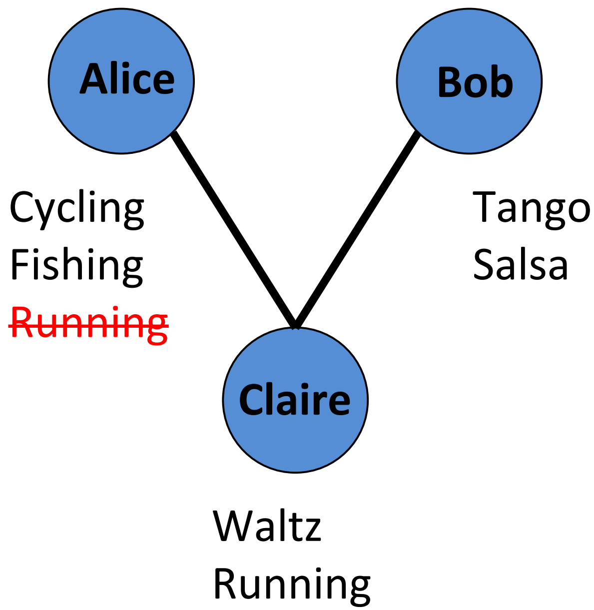
- Models networks as they over time, by way of changing latent features



J. R. Foulds, A. Asuncion, C. DuBois, C. T. Butts, P. Smyth.
A dynamic relational infinite feature model for longitudinal social networks. AISTATS 2011

Dynamic Relational Infinite Feature Model (DRIFT)

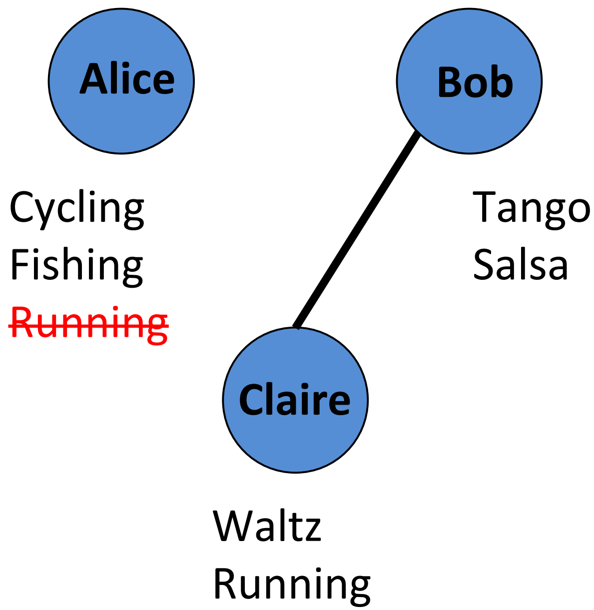
- Models networks as they over time, by way of changing latent features



J. R. Foulds, A. Asuncion, C. DuBois, C. T. Butts, P. Smyth.
A dynamic relational infinite feature model for longitudinal social networks. AISTATS 2011

Dynamic Relational Infinite Feature Model (DRIFT)

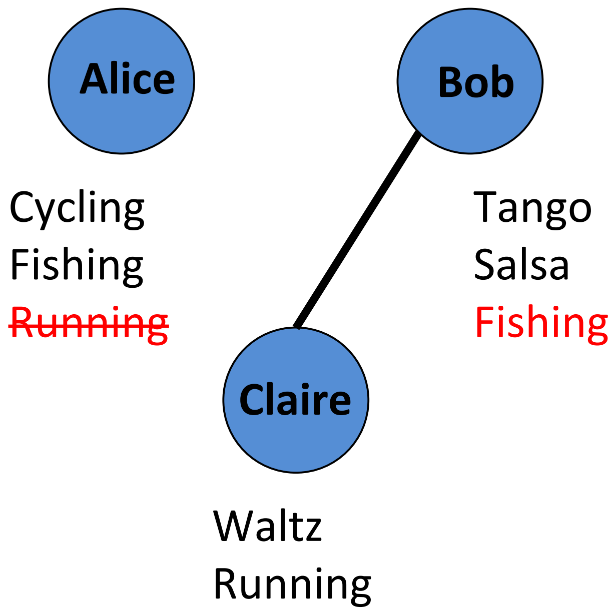
- Models networks as they over time, by way of changing latent features



J. R. Foulds, A. Asuncion, C. DuBois, C. T. Butts, P. Smyth.
A dynamic relational infinite feature model for longitudinal social networks. AISTATS 2011

Dynamic Relational Infinite Feature Model (DRIFT)

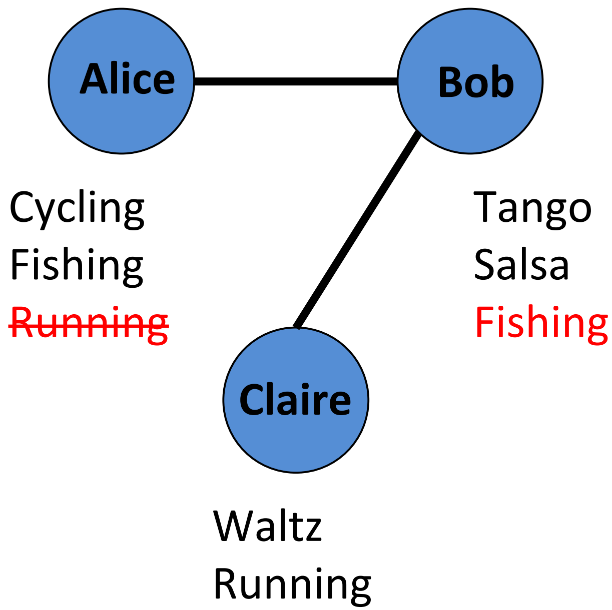
- Models networks as they over time, by way of changing latent features



J. R. Foulds, A. Asuncion, C. DuBois, C. T. Butts, P. Smyth.
A dynamic relational infinite feature model for longitudinal social networks. AISTATS 2011

Dynamic Relational Infinite Feature Model (DRIFT)

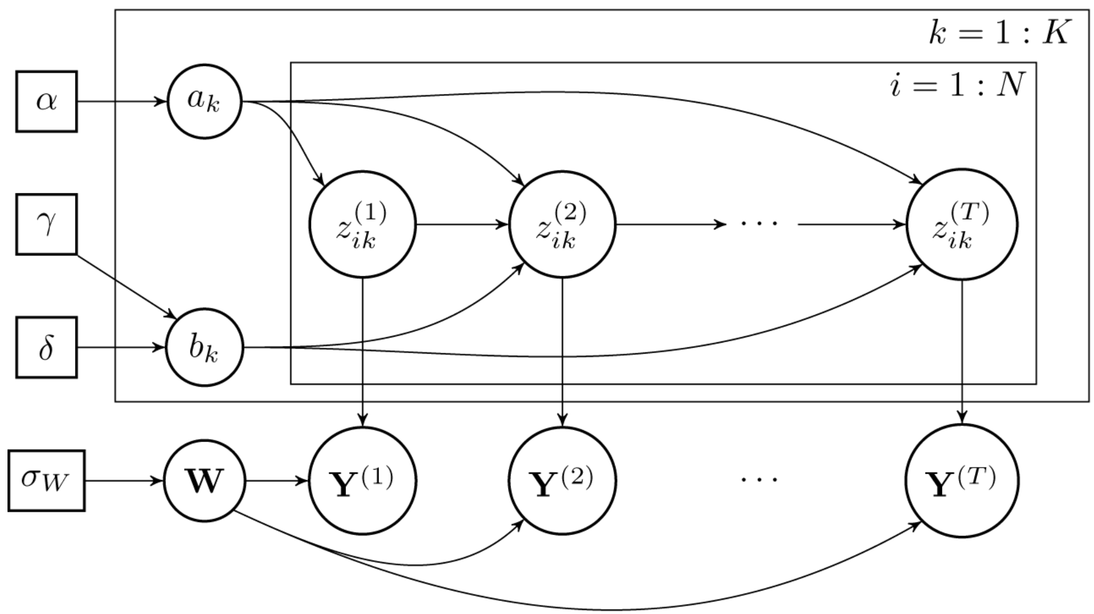
- Models networks as they over time, by way of changing latent features



J. R. Foulds, A. Asuncion, C. DuBois, C. T. Butts, P. Smyth.
A dynamic relational infinite feature model for longitudinal social networks. AISTATS 2011

Dynamic Relational Infinite Feature Model (DRIFT)

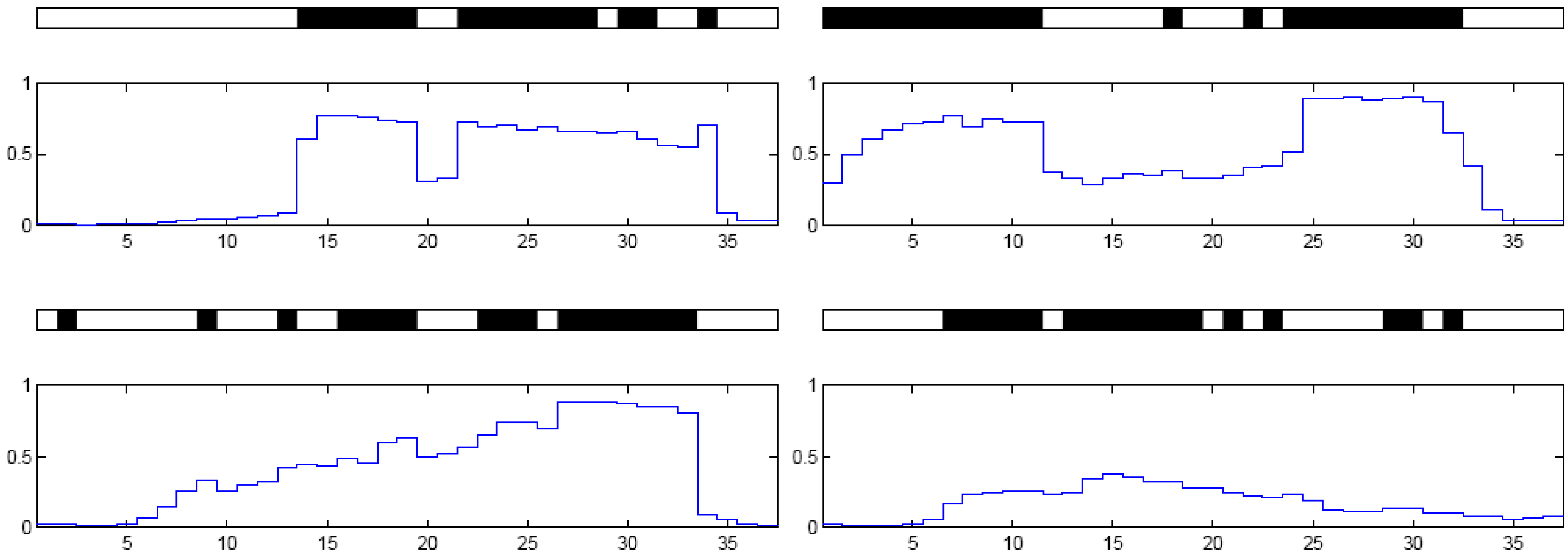
- Models networks as they over time, by way of changing latent features



- HMM dynamics for each actor/feature (factorial HMM)

J. R. Foulds, A. Asuncion, C. DuBois, C. T. Butts, P. Smyth.
A dynamic relational infinite feature model for longitudinal social networks. AISTATS 2011

Enron Email Data: Edge Probability Over Time



J. R. Foulds, A. Asuncion, C. DuBois, C. T. Butts, P. Smyth.
A dynamic relational infinite feature model for longitudinal social networks. AISTATS 2011

Quantitative Results

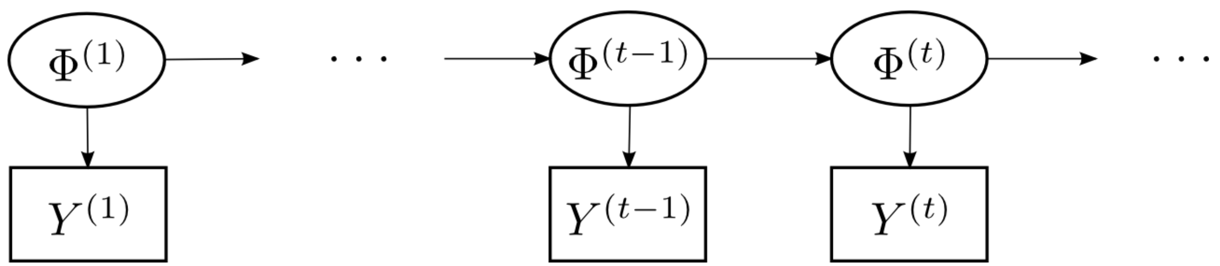
Synthetic Dataset	Naive	Baseline	LFRM (last/current)	LFRM (all)	DRIFT
Forecast LL	-31.6	-32.6	-28.4	-31.6	**−11.6**
Missing Data LL	-575	-490	-533	-478	**−219**
Forecast AUC	N/A	0.608	0.779	0.596	**0.939**
Missing Data AUC	N/A	0.689	0.675	0.691	**0.925**

Enron Dataset	Naive	Baseline	LFRM (last/current)	LFRM (all)	DRIFT
Forecast LL	-141	-108	-119	-98.3	**−83.5**
Missing Data LL	-1610	-1020	-1410	-981	**−639**
Forecast AUC	N/A	0.874	0.777	0.891	**0.910**
Missing Data AUC	N/A	0.921	0.803	0.933	**0.979**

J. R. Foulds, A. Asuncion, C. DuBois, C. T. Butts, P. Smyth.
A dynamic relational infinite feature model for longitudinal social networks. AISTATS 2011

Hidden Markov dynamic network models

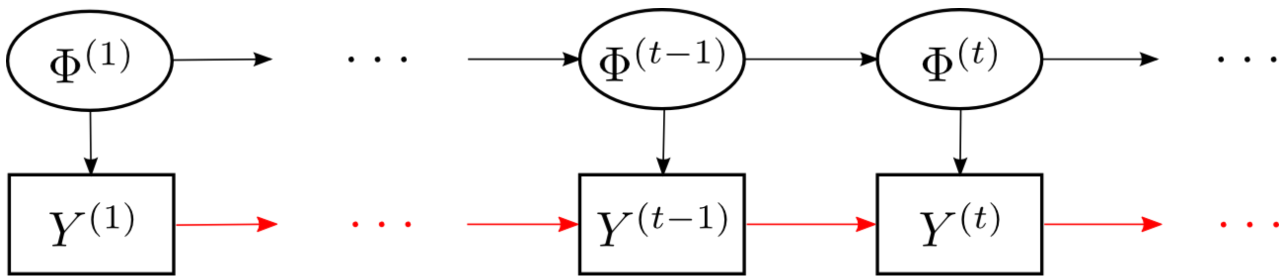
- Most work on dynamic network modeling assumes hidden Markov structure
  - Latent variables and/or parameters follow Markov dynamics
  - Graph snapshot at each time generated using static network model, e.g. stochastic block model or latent feature model as in DRIFT



  - Has been used to extend SBMs to dynamic models (Yang et al., 2011; Xu and Hero, 2014)

Beyond hidden Markov networks

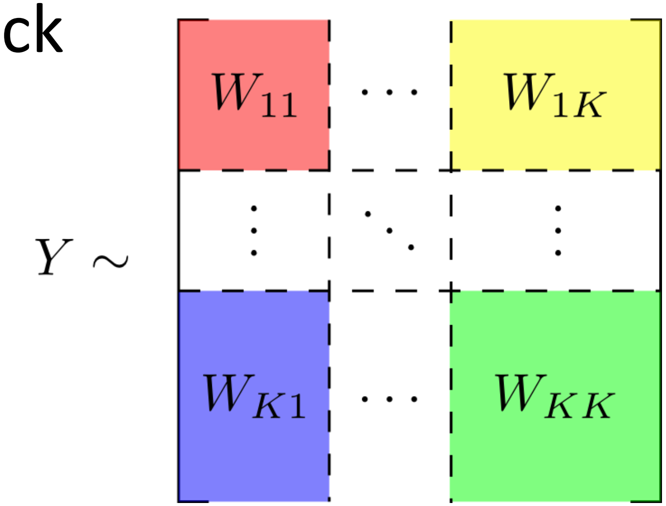
- Hidden Markov model (HMM) structure is tractable but not very realistic assumption in social interaction networks
  - Interaction between two people does not influence future interactions
- Autoregressive HMM: Allow current graph to depend on current parameters and previous graph



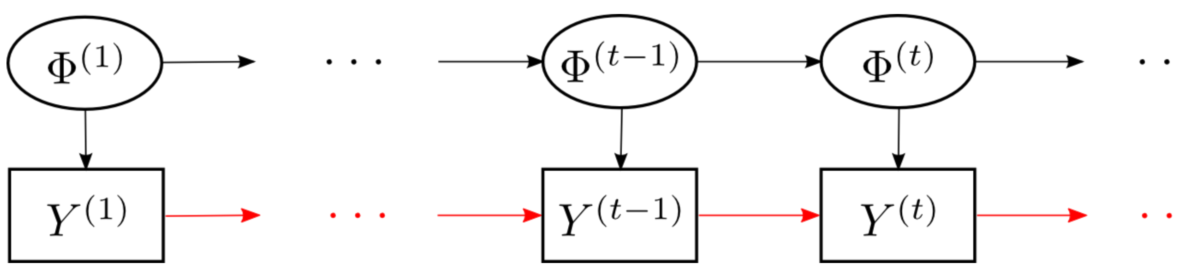
- Approximate inference using extended Kalman filter + greedy algorithms
  - Scales to ~ 1000 nodes

Stochastic block transition model

- Main idea: parameterize each block (k, k') with two probabilities
  - Probability of forming new edge
    $\pi_{kk'}^{t|0} = \Pr\left(Y_{ij}^{(t)} = 1 | Y_{ij}^{(t-1)} = 0\right)$
  - Probability of existing edge re-occurring
    $\pi_{kk'}^{t|1} = \Pr\left(Y_{ij}^{(t)} = 1 | Y_{ij}^{(t-1)} = 1\right)$

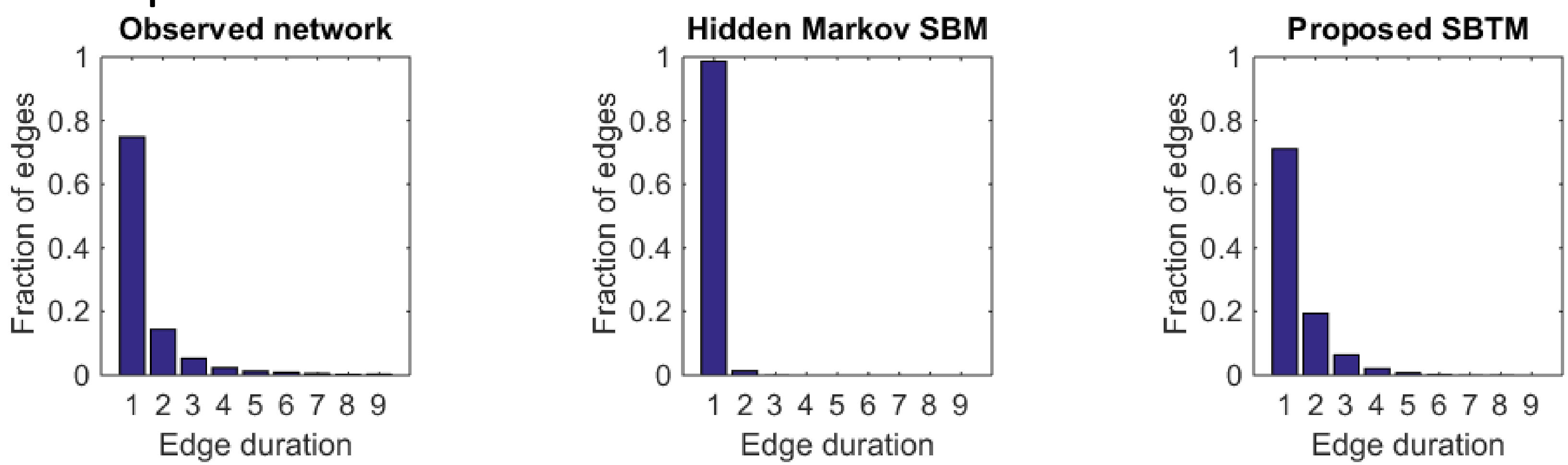


- Generate graph at initial time step using SBM
- Place Markov model on $\Pi^{t|0}, \Pi^{t|1}$



Application to Facebook wall posts

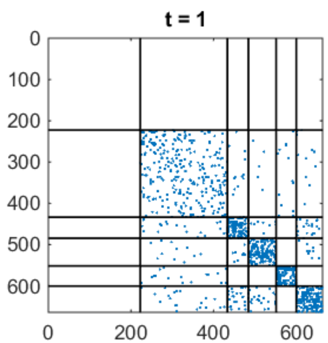
- Fit dynamic SBMs to network of Facebook wall posts
  - ~ 700 nodes, 9 time steps, 5 classes
- How accurately do hidden Markov SBM and SBTM replicate edge durations in observed network?
  - Simulate networks from both models using estimated parameters

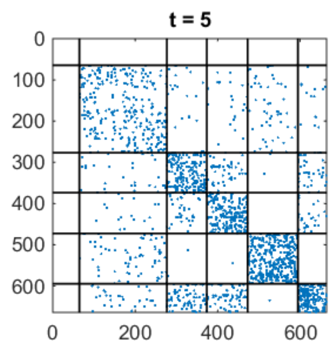


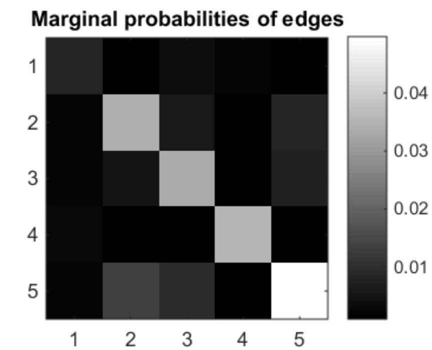
  - Hidden Markov SBM cannot replicate long-lasting edges in sparse blocks

Behaviors of different classes

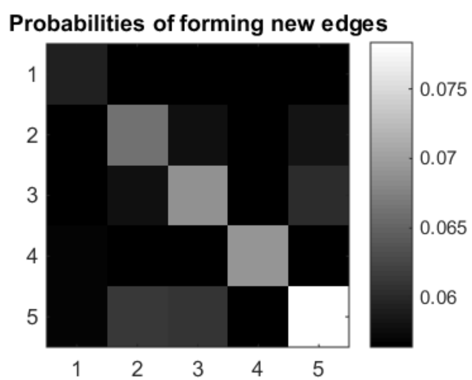
- SBTM retains interpretability of SBM at each time step

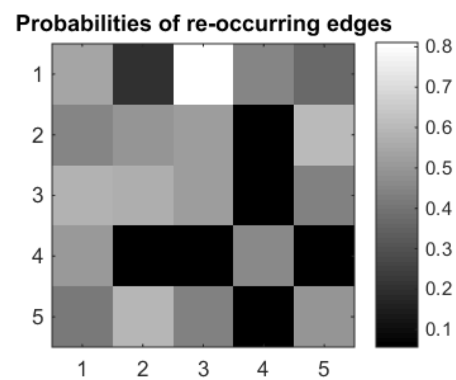






- Q: Do different classes behave differently in how they form edges?





- A: Only for probability of existing edges re-occurring
- New insight revealed by having separate probabilities in SBTM

Summary

- Generative models provide a powerful mechanism for modeling and analyzing social media data
- Latent variable models offer flexible yet interpretable models motivated by sociological principles
  - Latent space model
  - Stochastic block model
  - Mixed-membership stochastic block model
  - Latent feature model
- Generative models provide a rich mechanism for incorporating multiple modalities of data present in social media
  - Dynamic networks, cascades, joint modeling with text