Specifying node characteristics by combining social network data and user-generated -content

Mirai Igarashi (D2)
Nobuhiko Terui (Co-author)

You can download this slide from https://igarashim.github.io

1. Introduction

- Goal for marketing is … “to know the personality”
  - Traditionally, we use demographics, questionnaire and purchase behavior data
  - But people form their social network and post some contents to represent characteristics
  - Such unstructured data is hard to analyze, but rich information
- Goal of our study is to propose a **statistical model for estimating “characteristics”** by combining social network and user-generated-content

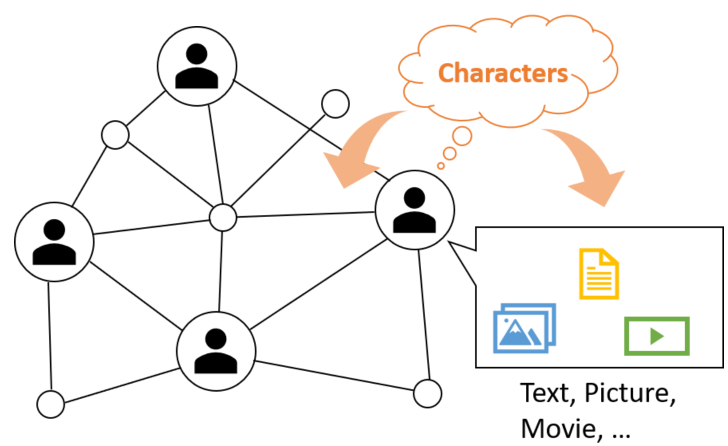
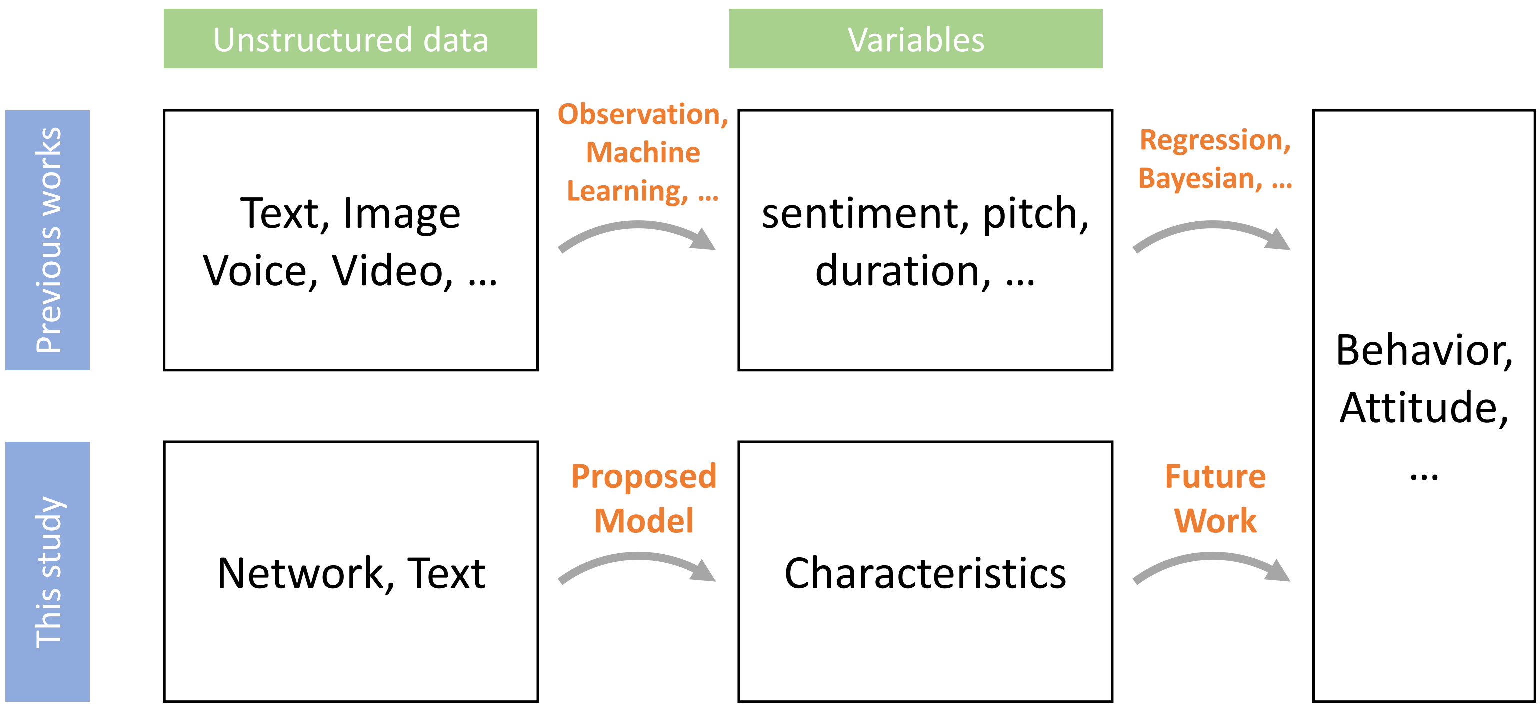


Table of contents

2. Literature review

- Many marketing researchers use unstructured data
  - ➢ Use some measurements for unstructured data (Text semantics, Ludwig et al 2013; Facial and gestural cues, Singh et al 2018)
  - ➢ Our model measures people's characteristics



2. Literature review

Some topic models for network and text data are proposed

- For **network**
  - ➢ Stochastic block model (SBM, Wang & Wong 1987)
  - ➢ Mixed Membership Stochastic blockmodel (MMSB, Airoldi et al 2008)
- For **text**
  - ➢ Latent Dirichlet allocation (LDA, Blei et al. 2003)
  - ➢ Recently, developed in marketing (e.g. Tirunillai & Tellis 2014; Toubia et al 2019)
- For **network and text**
  - ➢ Community user topic model (CUT, Zhou et al 2006)
  - ➢ Community author recipient topic-model (CART, Pathak et al 2006)
  - ➢ Topic-link LDA (TL-LDA, Liu et al 2009)
  - ➢ Stochastic topic block model (STBM, Bouveyron et al 2018)

3.1 Model specification

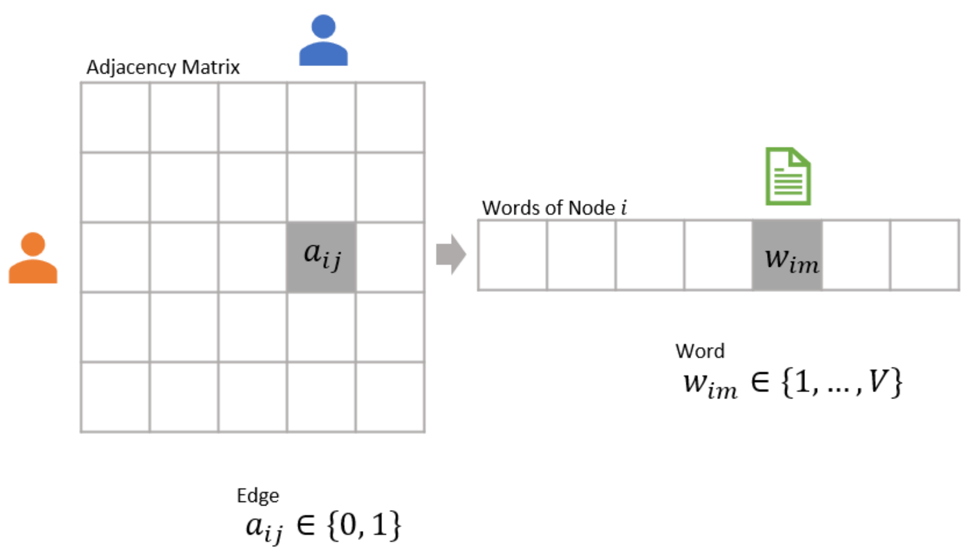
Data

- Adjacency matrix A (0: not connected, 1: connected)

$$a_{ij} \in \{0, 1\}, \qquad i, j = 1, \ldots, D$$

- Bag of words W (1: baseball, 2: book, …, V: iPhone)

$$w_{im} \in \{1, \ldots, V\}, \qquad m = 1, \ldots, M_i$$

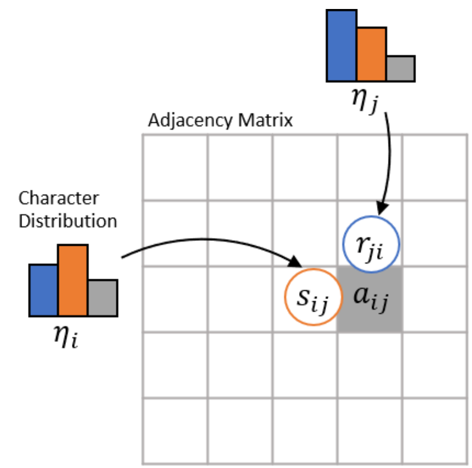


3.1 Model specification

Network

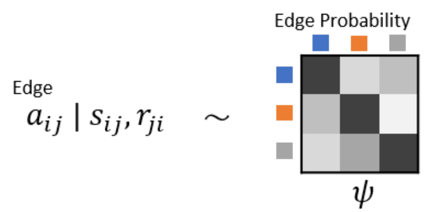
- For the edge $i \rightarrow j$, sender i and recipient j have latent characteristics (s_{ij}, r_{ji}) according to **character distribution** (η)

$$s_{ij} \sim Categorical(\eta_i), \qquad r_{ji} \sim Categorical(\eta_j)$$
$$s_{ij}, r_{ji} \in \{1, \ldots, K\}$$
$$\sum_{k=1}^{K} \eta_{ik} = 1, \ \forall i, \qquad \eta_{ik} \geq 0, \ \forall k$$



- When s_{ij}, r_{ji} are given, edge a_{ij} is generated according to **edge probability** (ψ)

$$a_{ij} | s_{ij}, r_{ji} \sim Bernoulli(\psi_{s_{ij}, r_{ji}}),$$
$$0 \leq \psi_{kk'} \leq 1, \ \forall k, k'$$



3.1 Model specification

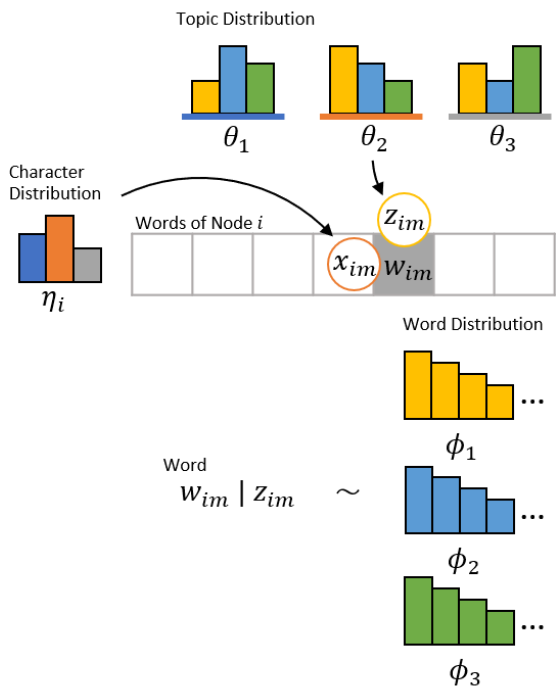
Text

- Node i's m-th word has latent characteristic (x_{im}) and latent topic (z_{im}) according to character distribution (η) and **topic distribution** (θ)

$$x_{im} \sim Categorical(\eta_i), \qquad x_{im} \in \{1, \ldots, K\}$$
$$z_{im}|x_{im} \sim Categorical(\theta_{x_{im}}), \qquad z_{im} \in \{1, \ldots, L\}$$
$$\sum_{l=1}^{L} \theta_{kl}, \ \forall k, \qquad \theta_{kl} \geq 0, \ \forall l$$

- When z_{im} is given, word w_{im} is generated according to **word distribution** (ϕ)

$$w_{im}|z_{im} \sim Categorical(\phi_{z_{im}})$$
$$\sum_{v=1}^{V} \phi_{lv}, \ \forall l, \qquad \phi_{lv} \geq 0, \ \forall v$$



3.2 Estimation

We set prior distributions according to the conjugacy

Likelihood	Prior distribution	(Full conditional) Posterior distribution
$P(s_{ij}\|\eta_i) = Categorical(\eta_i)$ $P(r_{ij}\|\eta_i) = Categorical(\eta_i)$ $P(x_{im}\|\eta_i) = Categorical(\eta_i)$	$P(\eta_i\|\gamma) = Dirichlet(\gamma)$	$P(\eta_i\|s_i, r_i, x_i, \gamma) =$ $Dirichlet(N_i + M_i + \gamma_k)$
$P(a_{ij}\|s_{ij}, r_{ji}, \psi) =$ $Bernoulli(\psi_{s_{ij}, r_{ji}})$	$P(\psi_{kk'}\|\delta, \epsilon) = Beta(\delta, \epsilon)$	$P(\psi_{kk'}\|A, S, R, \delta, \epsilon) =$ $Beta(n^{(p)}_{kk'} + \delta, n^{(m)}_{kk'} + \epsilon)$
$P(z_{im}\|x_{im}, \theta) =$ $Categorical(\theta_{x_{im}})$	$P(\theta_k\|\alpha) = Dirichlet(\alpha)$	$P(\theta_k\|X, Z, \alpha) =$ $Dirichlet(M_k + \alpha)$
$P(w_{im}\|z_{im}, \phi) =$ $Categorical(\phi_{z_{im}})$	$P(\phi_l\|\beta) = Dirichlet(\beta)$	$P(\phi_l\|W, Z, \beta) =$ $Dirichlet(M_l + \beta)$

3.2 Estimation

We use collapsed Gibbs sampling for estimating parameters

By integrating out parameters, conditional posterior distribution of latent variables can be derived as follows (Igarashi & Terui 2019):

$$P(s_{ij}=k, r_{ji}=k' | a_{ij}, A_{\backslash ij}, S_{\backslash ij}, R_{\backslash ji}, X, \gamma, \delta, \epsilon)$$
$$= \frac{N_{ik\backslash ij} + M_{ik} + \gamma_k}{\sum_t \left(N_{it\backslash ij} + M_{it} + \gamma_t\right)} \times \frac{N_{jk'\backslash ji} + M_{jk'} + \gamma_{k'}}{\sum_t \left(N_{jt\backslash ji} + M_{jt} + \gamma_t\right)} \times \frac{\left(n^{(p)}_{kk'\backslash ij} + \delta_{kk'}\right)^{\mathbb{I}(a_{ij}=1)} \left(n^{(m)}_{kk'\backslash ij} + \epsilon_{kk'}\right)^{\mathbb{I}(a_{ij}=0)}}{n^{(p)}_{kk'\backslash ij} + n^{(m)}_{kk'\backslash ij} + \delta_{kk'} + \epsilon_{kk'}}$$

$$P(x_{im}=k, z_{im}=l | w_{im}=v, W_{\backslash im}, S, R, X_{\backslash im}, Z_{\backslash im}, \alpha, \beta, \gamma)$$
$$= \frac{M_{lv\backslash im} + \beta_v}{\sum_u \left(M_{lu\backslash im} + \beta_u\right)} \times \frac{M_{kl\backslash im} + \alpha_l}{\sum_q \left(M_{kq\backslash im} + \alpha_q\right)} \times \frac{N_{ik} + M_{ik\backslash im} + \gamma_k}{\sum_t \left(N_{it} + M_{it\backslash im} + \gamma_t\right)}$$

3.2 Estimation

Using Gibbs samples from above equations, parameters are point-estimated as follows:

$$\hat{\eta}_{ik} = \frac{1}{G-b}\sum_{g=b+1}^{G}\frac{N_{ik}^{(g)} + M_{ik}^{(g)} + \gamma_k}{\sum_t \left(N_{it}^{(g)} + M_{it}^{(g)} + \gamma_t\right)}$$

$$\hat{\psi}_{kk'} = \frac{1}{G-b}\sum_{g=b+1}^{G}\frac{n_{kk'}^{(p,g)} + \delta_{kk'}}{n_{kk'}^{(p,g)} + n_{kk'}^{(m,g)} + \delta_{kk'} + \epsilon_{kk'}}$$

$$\hat{\theta}_{kl} = \frac{1}{G-b}\sum_{g=b+1}^{G}\frac{M_{kl}^{(g)} + \alpha_l}{\sum_q \left(M_{kl}^{(g)} + \alpha_q\right)}$$

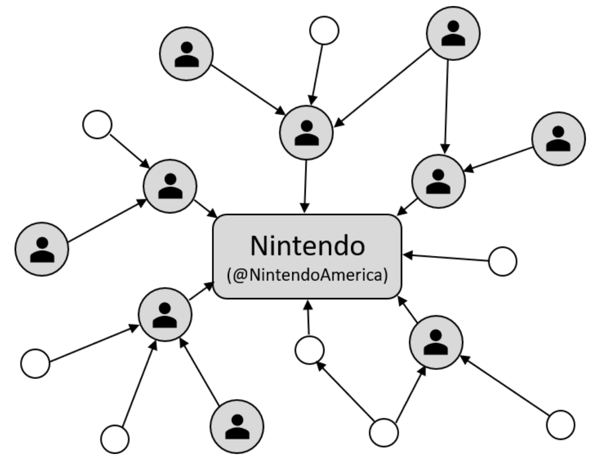
$$\hat{\phi}_{lv} = \frac{1}{G-b}\sum_{g=b+1}^{G}\frac{M_{lv}^{(g)} + \beta_v}{\sum_u \left(M_{lu}^{(g)} + \beta_u\right)}.$$

Numbers of characteristics (K) and topics (L) is determined by grid search using WAIC (Watanabe 2010)

4.1 Dataset

Twitter's network and text collected by authors consist of:

- Ego-network centered on Nintendo account (following relationship at May 1, 2018)
- Users' tweets posted on their timeline (September 1, 2017 – February 28, 2018)



The summary of dataset after sampling and some preprocessing:

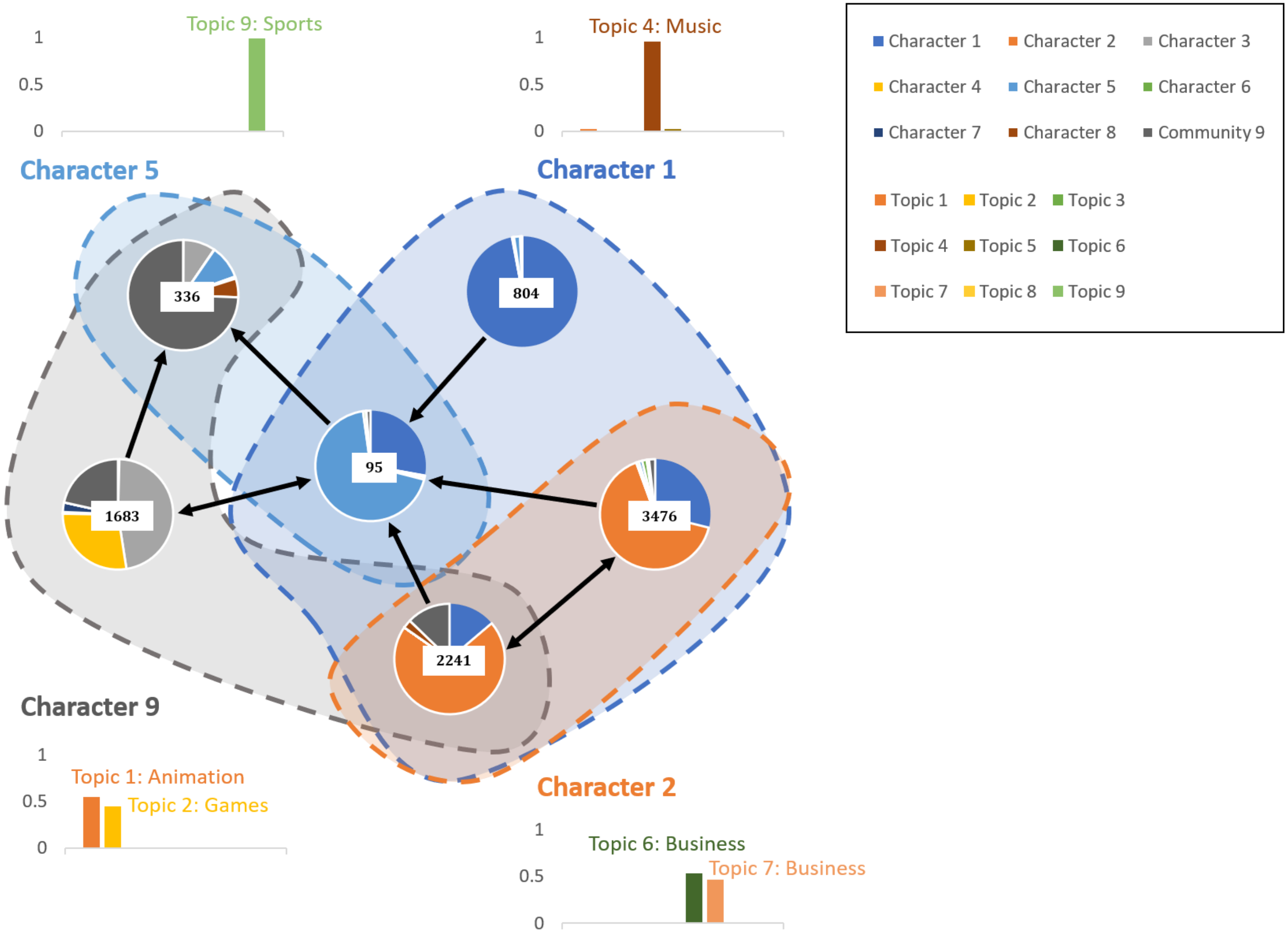
D (nodes)	V (words)	Ave. links (sparsity)	Ave. words (sparsity)
3,500	9,001	19.7 links (0.56%)	59.3 words (1.69%)

4.2 Empirical results

Top 10 words frequently appearing in each topic

Topic 1 (Animation)	Topic 2 (Game)	Topic 3 (E-sports)	Topic 4 (Music)	Topic 5 (Every life)	Topic 6 (Business)	Topic 7 (Business)	Topic 8 (Streaming Broadcasting)	Topic 9 (Sports)
podernfamili	vgc	hori	vevo	leed	trapadr	growthhack	nonfollow	zeldathon
gamedesign	savvi	mkleosaga	spinrilla	cto	digitalmarket	gdpr	teamemmmmsi	dokkan
criticalrol	gamedesign	wnf	lube	momlif	ddrive	socialmediamarket	twitchkitten	htgawm
blackclov	steinsgat	mdva	suav	dogsoftwitt	contentmarket	iartg	roku	orton
hunterxhunter	nyxl	hyrulesaga	drippin	beck	smm	smm	wizebot	oiler
jojosbizarreadventur	xenovers	cfl	ahscult	austria	amread	gainwithpyewaw	ryzen	sdlive
fursuitfriday	acnl	nood	wshh	hemp	bigdata	asmsg	airdrop	horford
tfc	artstat	qanba	ouija	tock	gdpr	ifb	dg	herewego
amiga	firer	zeku	foodporn	crowdfir	gainwithxtiandela	digitalmarket	freebiefriday	rozier
sml	tamagotchi	junedecemb	sizzl	monaco	fiverr	css	streamersconnect	earnhistori

4.2 Empirical results



4.3 Prediction

- Settings

Network (19.7 edges/node) $\xrightarrow{10\%}$ Test samples

Network (19.7 edges/node) $\xrightarrow{90\%}$ Training samples

Text (59.3 words/node) $\xrightarrow{100\%}$ Training samples

- Prediction

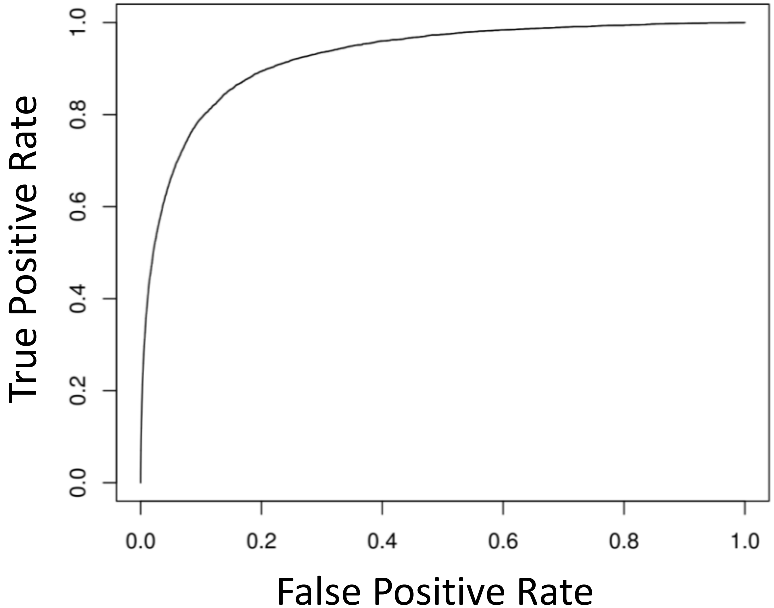
$$P(a_{ij} = 1) = \sum_{k=1}^{K} \sum_{k'=1}^{K} \widehat{\eta}_{ik} \cdot \widehat{\eta}_{jk'} \cdot \widehat{\psi}_{kk'}$$

- Evaluation
  - Area Under the Curve (AUC)
  - Matthews Correlation Coefficient (MCC)

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN)(FN + TN)(TN + FP)(FP + TP)}}$$

4.3 Prediction

- ROC curve & AUC



  - AUC: 0.93 (Perfect -> 1.00, At Random -> 0.5)
  - Our model has a good predictive performance

4.3 Prediction

- Confusion Matrix

		Prediction	
		Link	Non-link
Data	Link	2,041	4,786
	Non-link	7,079	1,211,094

  - Cutoff: 0.08
  - MCC: 0.254
  - True Positive Rate: $\frac{2{,}041}{2{,}041+4{,}786} \approx 29.9\%$

- Our model
  - predicts link edges with adequate accuracy
  - has a room for improvement on predicting task (ex. modeling sparsity, Airoldi et al 2008; Latouche et al 2011)

Conclusion (STATIC)

- We estimate **people's characteristics** from **network and text** data on social media
  - ➢ we propose a new topic model
- People generate social networks and text contents according to their characteristics
  - ➢ character distribution (η_i) reflects the characters
- The proposed model demonstrated
  - ➢ estimating characteristics from network and text data
  - ➢ interpreting the characters by network information and word topics
  - ➢ predicting holdout edges with adequate accuracy

Future works

- Dealing with node heterogeneity by hierarchical structure
- Applying marketing model

1. Introduction

- Network and text change over time

  - "Character" and "topic" evolve over time

- Goal of this study is to extend our model to capture the **dynamics** of time-evolving network and text data

2. Literature review

Some topic models for network and text evolving over time

- For **network**
  - ➢ Dynamic Mixed Membership SBM (dMMSB, Xing et al 2010)
  - ➢ Dynamic SBM (dSBM, Yang et al 2011)
- For **text**
  - ➢ Dynamic Topic model (DTM, Blei & Lafferty 2006)
  - ➢ Continuous time DTM (Wang, Blei, & Heckerman 2012)
- For **network and text**
  - ➢ Dynamic stochastic topic block model (dSTBM, Bouveyron et al 2019)

3.1 Model specification

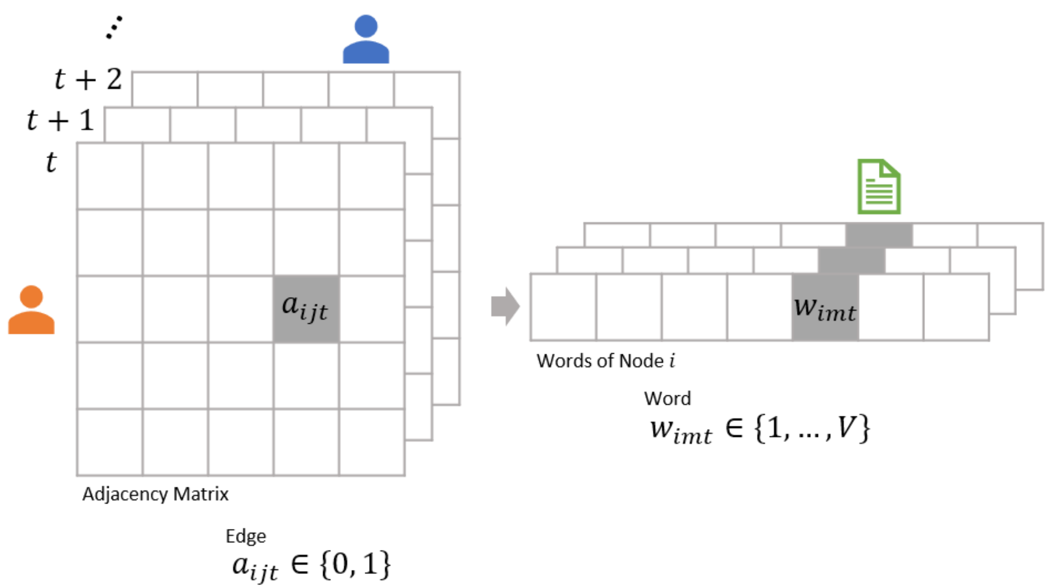
Data

- Adjacency matrix A (0: not connected, 1: connected)

$$a_{ijt} \in \{0, 1\}, \qquad i, j = 1, \ldots, D, \quad t = 1, \ldots, T$$

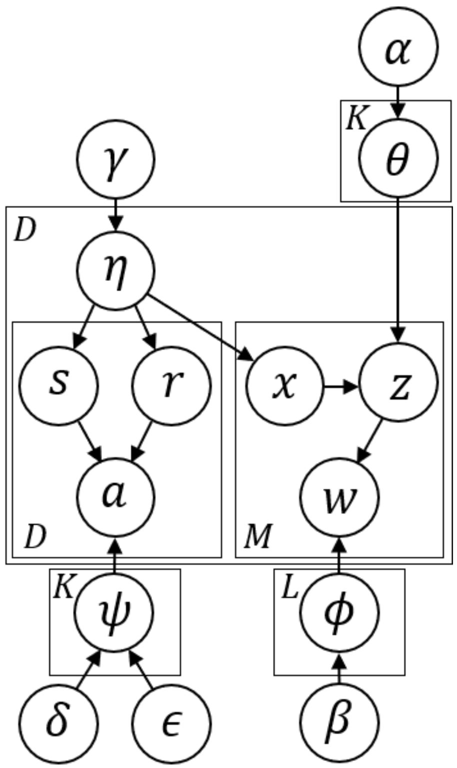
- Bag of words W (1: baseball, 2: book, …, V: iPhone)

$$w_{imt} \in \{1, \ldots, V\}, \qquad m = 1, \ldots, M_{it}, \quad t = 1, \ldots, T$$

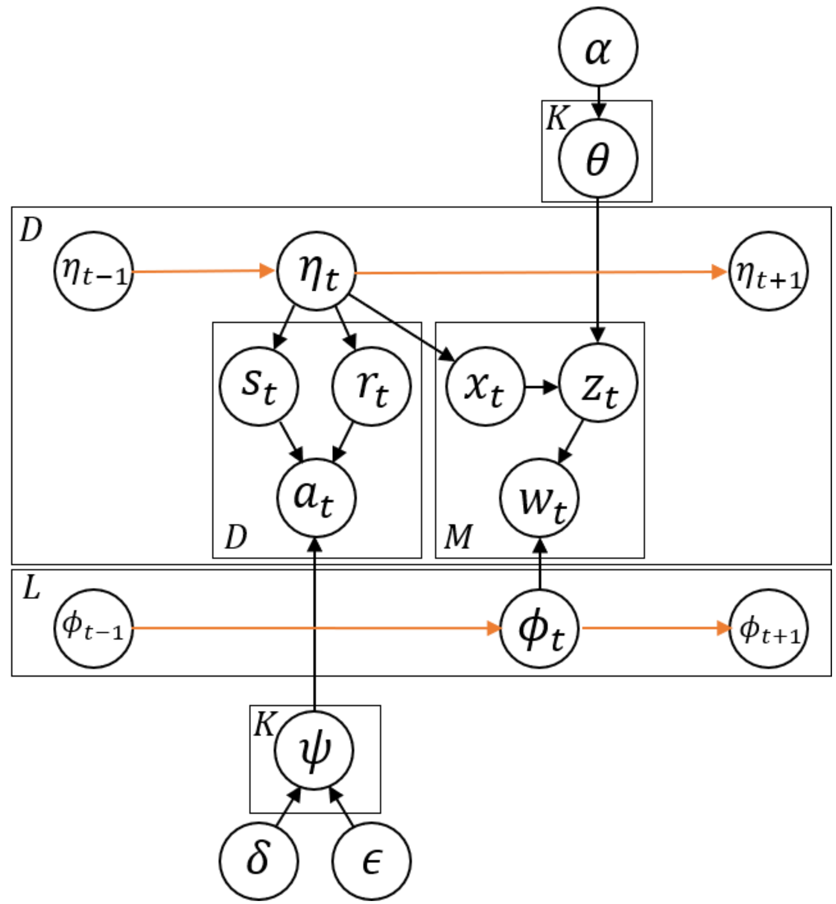


3.1 Model specification

Static proposed model



Dynamic proposed model



3.1 Model specification

Network

- **Character distribution** (η_t) evolves with Gaussian noise

$$\eta_{it}|\eta_{it-1} \sim N_K\left(\eta_{it-1}, \sigma_\eta^2 I\right)$$

- For the edge $i \rightarrow j$ at time t, sender i and recipient j have latent characteristics (s_{ijt}, r_{jit}) according to normalized character distribution

$$s_{ijt} \sim Categorical\left(\pi\left(\eta_{it}\right)\right), \qquad r_{jit} \sim Categorical\left(\pi\left(\eta_{jt}\right)\right)$$

$$\pi(x) = \frac{exp(x_k)}{\sum_{k'} exp(x_{k'})}$$

- When s_{ijt}, r_{jit} are given, edge a_{ijt} is generated according to **edge probability** (ψ)

$$a_{ijt}|s_{ijt}, r_{jit} \sim Bernoulli\left(\psi_{s_{ijt}, r_{jit}}\right)$$

3.1 Model specification

Text

- Node i's m-th word at time t has latent characteristic (x_{imt}) and latent topic (z_{imt}) according to character distribution (η_t) and **topic distribution** (θ)

$$x_{imt} \sim Categorical(\pi\left(\eta_{it}\right)), \qquad z_{imt} | x_{imt} \sim Categorical(\theta_{x_{imt}})$$

- **word distribution** (ϕ_t) evolves with Gaussian noise

$$\phi_{lt} | \phi_{lt-1} \sim N_V\left(\phi_{lt-1}, \sigma_\phi^2 I\right)$$

- When z_{imt} is given, word w_{imt} is generated according to normalized word distribution

$$w_{imt} | z_{imt} \sim Categorical\left(\pi\left(\phi_{z_{imt} t}\right)\right)$$

3.2 Estimation

- In static model, we set conjugate prior distributions for parameters and derive (collapsed) Gibbs sampling.

  likelihood: $s_{ij}|\eta_i \sim Categorical\left(\eta_i\right)$ prior: $\eta_i \sim Dirichlet\left(\gamma\right)$

  posterior: $\eta_i|\cdot \sim Dirichlet\left(\cdot\right)$

- In dynamic model, prior distributions have no conjugacy.

  likelihood: $s_{ijt}|\eta_{it} \sim Categorical\left(\pi\left(\eta_{it}\right)\right)$ prior: $\eta_{it} \sim N\left(\eta_{it-1}, \sigma_\eta^2 I\right)$

  posterior: cannot be derived in the same form of prior

  -> **Variational Bayes**

- Variational Bayes explore variational posterior which is closest to true posterior in the sense of KL divergence.

$$\beta = \{\eta_{1:T}, \phi_{1:T}, \psi, \theta, s_{1:T}, r_{1:T}, x_{1:T}, z_{1:T}\}$$
$$q(\beta|data) = \arg\min_q \; KL[q(\beta)||p(\beta|data)]$$
$$\text{s.t. } q(\beta) \text{ is factorizable}$$

3.2 Estimation

Mean field assumption

$$q(\beta) = \prod_{i=1}^{D} \{q(\eta_{i1}, \dots, \eta_{iT})\} \times \prod_{l=1}^{L} \{q(\phi_{l1}, \dots, \phi_{lT})\} \times \prod_{k=1}^{K} \left\{ q(\theta_k) \prod_{k'=1}^{K} q(\psi_{kk'}) \right\}$$
$$\times \prod_{t=1}^{T} \left\{ \prod_{i=1}^{D} \left[\prod_{j=1}^{D} q(s_{ijt}) q(r_{jit}) \right] \left[\prod_{m=1}^{M_{it}} q(x_{imt}) q(z_{imt}) \right] \right\}$$

- $q(\eta_{i1}, \dots, \eta_{iT})$ and $q(\phi_{l1}, \dots, \phi_{lT})$ should not be factorized any more because these joint distributions have time dependence.

Variational Bayes + **Kalman filter** can be used (Blei & Lafferty 2006).

Model

$$\begin{cases} \eta_{it} | \eta_{it-1} \sim N(\eta_{it-1}, \sigma_\eta^2 I) \\ s_{ijt} | \eta_{it} \sim Categorical\left(\pi(\eta_{it})\right) \end{cases}$$

Approximation

$$\begin{cases} \eta_{it} | \eta_{it-1} \sim N(\eta_{it-1}, \sigma_\eta^2 I) \\ \hat{\eta}_{it} | \eta_{it} \sim N(\eta_{it}, \rho_\eta^2 I) \end{cases}$$

Variational observations

- Kalman filter can be the closed form Bayesian solution to the linear Gaussian filtering problem.
- Estimation of $\eta_{1:T}$ by Kalman filter and estimation of $\hat{\eta}_{1:T}$ by VB.

3.2 Estimation

Filtering distribution

$$q(\eta_{it}|\hat{\eta}_{i1:t}) = N\left(\mu_{it}, \lambda_{it}^2 I\right)$$

$$\mu_{it} = \left(\frac{\rho_\eta^2}{\lambda_{it}^2 + \sigma_\eta^2 + \rho_\eta^2}\right)\mu_{it-1} + \left(1 - \frac{\rho_\eta^2}{\lambda_{it}^2 + \sigma_\eta^2 + \rho_\eta^2}\right)\hat{\eta}_{it}, \quad \lambda_{it}^2 = \left(\frac{\rho_\eta^2}{\lambda_{it}^2 + \sigma_\eta^2 + \rho_\eta^2}\right)(\lambda_{it-1}^2 + \sigma_\eta^2)$$

$$q(\phi_{lt}|\hat{\phi}_{l1:t}) = N\left(\pi_{lt}, \omega_{lt}^2 I\right)$$

$$\pi_{lt} = \left(\frac{\rho_\phi^2}{\omega_{lt}^2 + \sigma_\phi^2 + \rho_\phi^2}\right)\pi_{lt-1} + \left(1 - \frac{\rho_\phi^2}{\omega_{lt}^2 + \sigma_\phi^2 + \rho_\phi^2}\right)\hat{\phi}_{lt}, \quad \omega_{lt}^2 = \left(\frac{\rho_\phi^2}{\omega_{lt}^2 + \sigma_\phi^2 + \rho_\phi^2}\right)(\omega_{lt-1}^2 + \sigma_\phi^2)$$

Smoothing distribution

$$q(\eta_{it}|\hat{\eta}_{i1:T}) = N\left(\tilde{\mu}_{it}, \tilde{\lambda}_{it} I\right)$$

$$\tilde{\mu}_{it} = \left(1 - \frac{\lambda_{it}^2}{\lambda_{it}^2 + \sigma_\eta^2}\right)\mu_{it} + \left(\frac{\lambda_{it}^2}{\lambda_{it}^2 + \sigma_\eta^2}\right)\tilde{\mu}_{it+1}, \quad \tilde{\lambda}_{it}^2 = \lambda_{it}^2 + \left(\frac{\lambda_{it}^2}{\lambda_{it}^2 + \sigma_\eta^2}\right)^2 \left(\tilde{\lambda}_{it+1}^2 - (\lambda_{it}^2 + \sigma_\eta^2)\right)$$

$$q(\phi_{lt}|\hat{\phi}_{l1:T}) = N\left(\tilde{\pi}_{lt}, \tilde{\omega}_{lt} I\right)$$

$$\tilde{\pi}_{lt} = \left(1 - \frac{\omega_{lt}^2}{\omega_{lt}^2 + \sigma_\phi^2}\right)\pi_{lt} + \left(\frac{\omega_{lt}^2}{\omega_{lt}^2 + \sigma_\phi^2}\right)\tilde{\pi}_{lt+1}, \quad \tilde{\omega}_{lt}^2 = \omega_{lt}^2 + \left(\frac{\omega_{lt}^2}{\omega_{lt}^2 + \sigma_\phi^2}\right)^2 (\tilde{\omega}_{lt+1}^2 - (\omega_{lt}^2 + \sigma_\phi^2))$$

3.2 Estimation

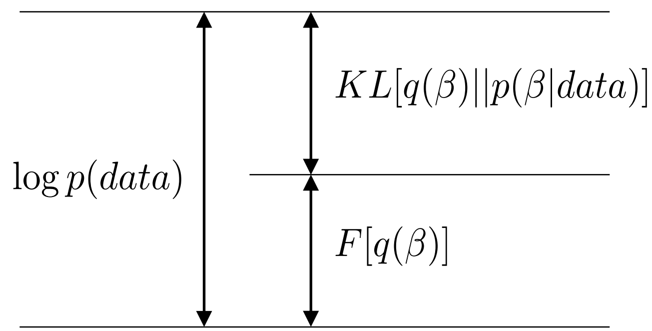
Evidence Lower Bound (ELBO)

- Minimizing KL divergence is equivalent to maximizing ELBO

$$\arg\min_q \; KL[q(\beta)||p(\beta|data)]$$

$$log(data) = F[q(\beta)] + KL[q(\beta)||p(\beta|data)]$$

$$F[q(\beta)] = \int q(\beta) \log \frac{p(data, \beta)}{q(\beta)} d\beta$$



$$\arg\max_q \; F[q(\beta)]$$

3.2 Estimation

Evidence Lower Bound (ELBO)

$$
\begin{aligned}
\log(a_{1:T}, w_{1:T}) \geq & \int \sum_{s,r,x,z} q(\beta_{1:T}) \log \frac{p(a_{1:T}, w_{1:T}, \eta_{1:T}, \phi_{1:T}, \psi, \theta)}{q(\eta_{1:T}, \phi_{1:T}, \psi, \theta)} d\eta d\phi d\psi d\theta \\
&= \mathbb{E}_{q(s,r,\psi)} \left[\log p(a_{1:T} | s_{1:T}, r_{1:T}, \psi)\right] + \mathbb{E}_{q(\psi)} \left[\log \frac{p(\psi)}{q(\psi)}\right] \\
&+ \mathbb{E}_{q(s,r,x,\eta)} \left[\log \frac{p(s_{1:T}|\eta_{1:T}) p(r_{1:T}|\eta_{1:T}) p(x_{1:T}|\eta_{1:T})}{q(s_{1:T}) q(r_{1:T}) q(x_{1:T})}\right] + \mathbb{E}_{q(\eta)} \left[\log \frac{p(\eta_{1:T})}{q(\eta_{1:T})}\right] \\
&+ \mathbb{E}_{q(z,\phi)} \left[\log p(w_{1:T} | z_{1:T}, \phi_{1:T})\right] + \mathbb{E}_{q(\phi)} \left[\log \frac{p(\phi_{1:T})}{q(\phi_{1:T})}\right] \\
&+ \mathbb{E}_{q(x,z,\theta)} \left[\log \frac{p(z_{1:T}|x_{1:T}, \theta)}{q(z_{1:T})}\right] + \mathbb{E}_{q(\theta)} \left[\log \frac{p(\theta)}{q(\theta)}\right]
\end{aligned}
$$

We find the stationary point by variation of ELBO for each parameter

-> **Update variational parameters** repeatedly until ELBO converged

We conduct **filtering and smoothing** for time evolving parameters (η_t, ϕ_t) using variational observations $(\hat{\eta}_t, \hat{\phi}_t)$

Conclusion (DYNAMIC)

- We propose dynamic topic model by combining time-evolving network and text contents.
- **Character distribution** (η_t) and **topic distribution** (ϕ_t) change over time with Gaussian noise.
- Estimation using combination of **variational Bayes** and **Kalman filter**
  - introduce **variational observations** to construct linear Gaussian state space model and obtain the exact solution by Kalman filtering and smoothing
  - update variational parameters at the stationary points

Future works

- Empirical analysis
- Apply for marketing model