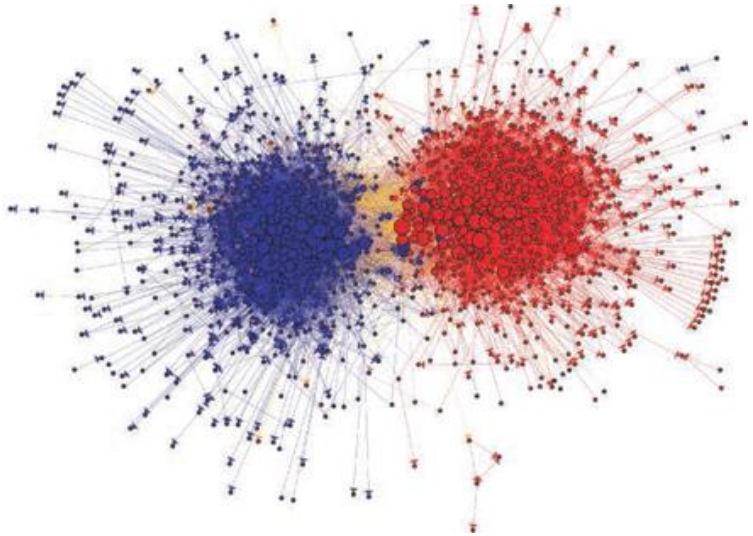
计算传播学研究：

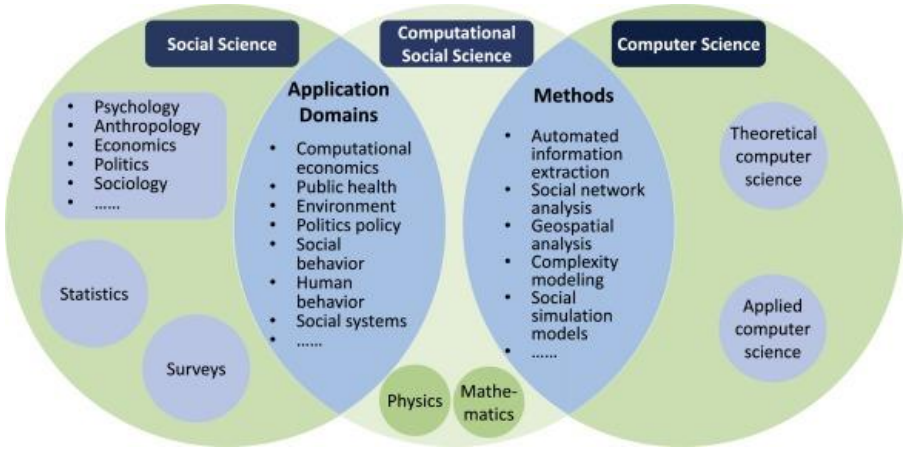
使用、结构、扩散、内容与效果

梁海 副教授
香港中文大学新闻与传播学院
hailiang@cuhk.edu.hk
https://drhailiang.com/

2023-04-03

什么是计算社会科学




Social Science
Computational Social Science
Computer Science
Application Domains
Methods
Psychology
Anthropology
Economics
Politics
Sociology
.....
Computational economics
Public health
Environment
Politics policy
Social behavior
Human behavior
Social systems
.....
Automated information extraction
Social network analysis
Geospatial analysis
Complexity modeling
Social simulation models
.....
Theoretical computer science
Applied computer science
Statistics
Surveys
Physics
Mathe-matics

什么是计算社会科学

- 经验的而非批判的
- 量化的而非质化的
- 数值的而非分析的

- 非介入的、自动的数据收集
- 自动的、数值的分析

什么是传播学

如何运用计算方法

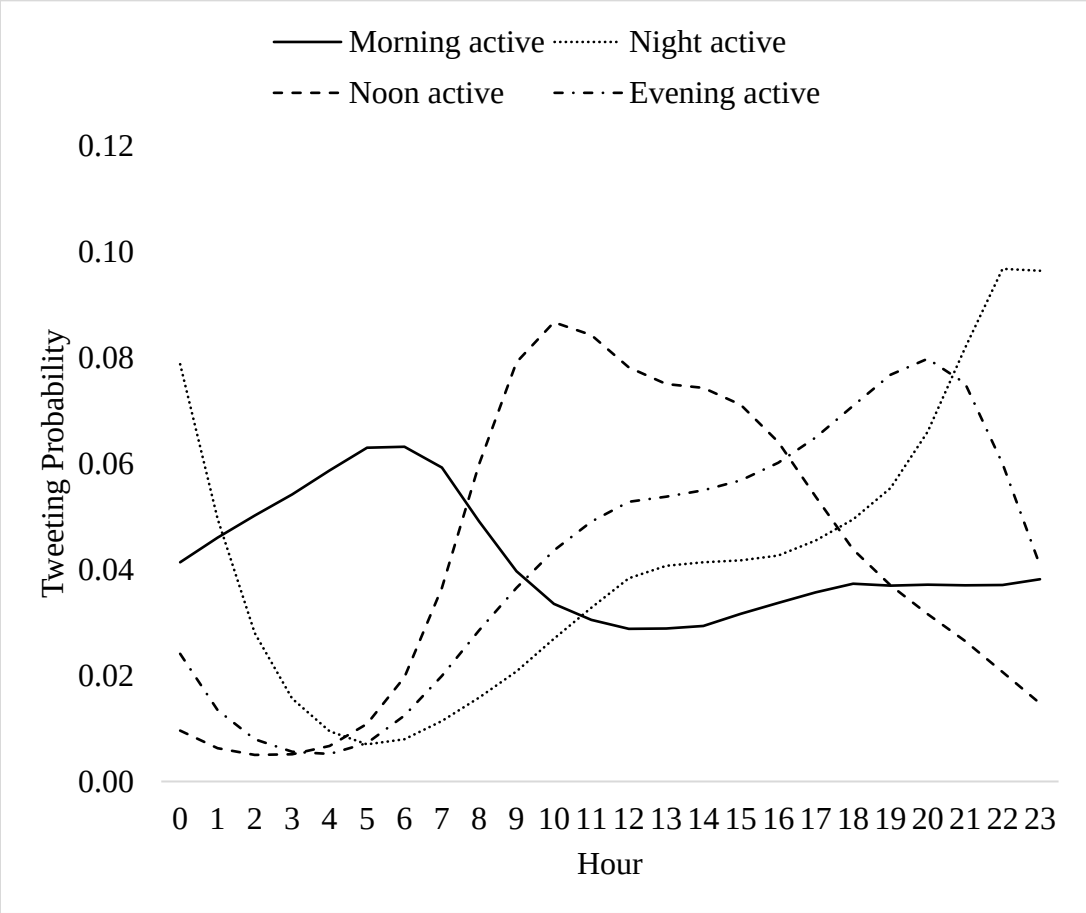
- 测量
  - 文本特征
  - 网络指标
- 描述
  - 虚假信息
  - 信息茧房
- 推断
  - 断续时间序列
  - 纵向网络分析
- 干预
  - 减少仇恨/极化
  - 用户界面设计

1. 使用 – 用户/行为分析

- 互联网使用
  - 采纳 (survey: use or not)
  - 行为模式
  - 生产/消费 (digital marketing/advertising)

- 受众/行为分析
  - 电视收视率
  - 动态 + 结构
  - 时间 + 网络

1. 使用 – 数字节律



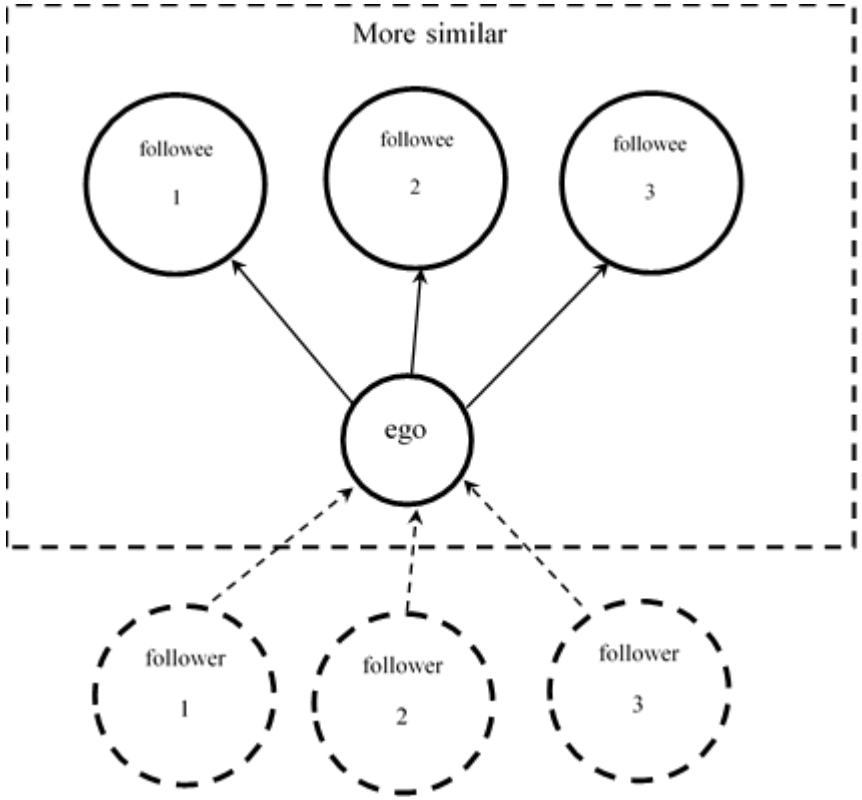
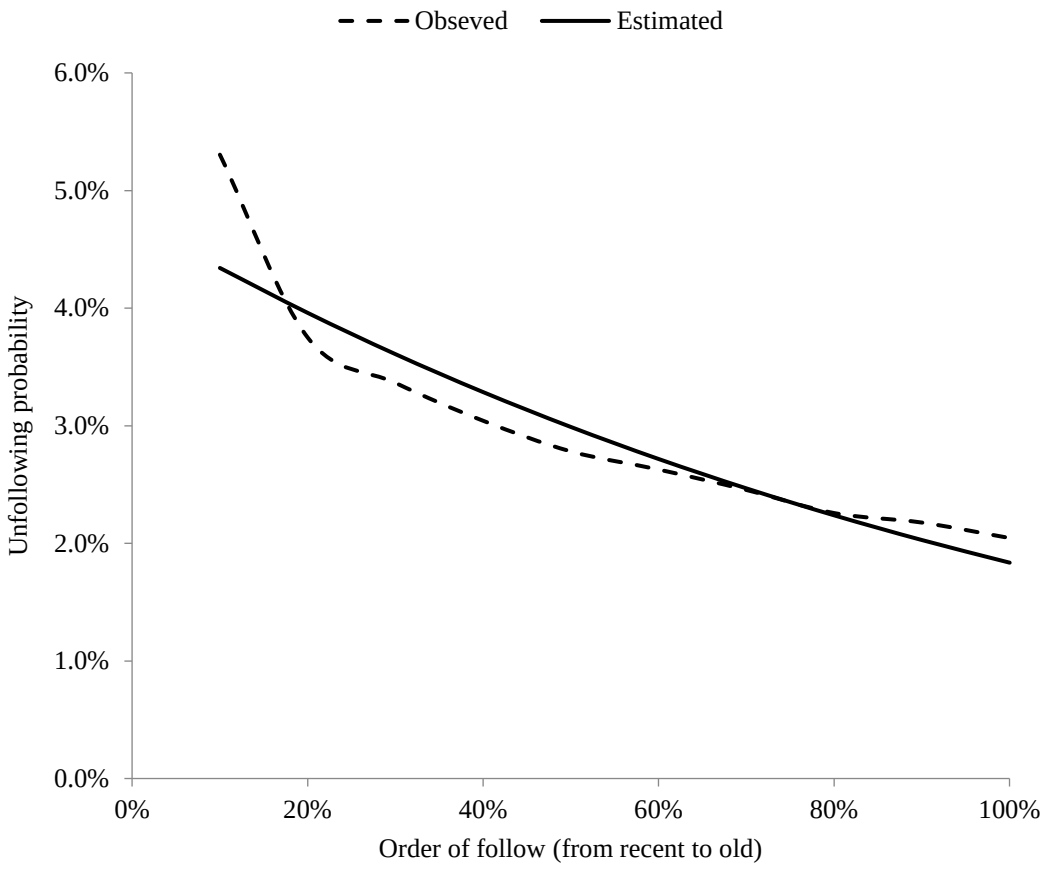


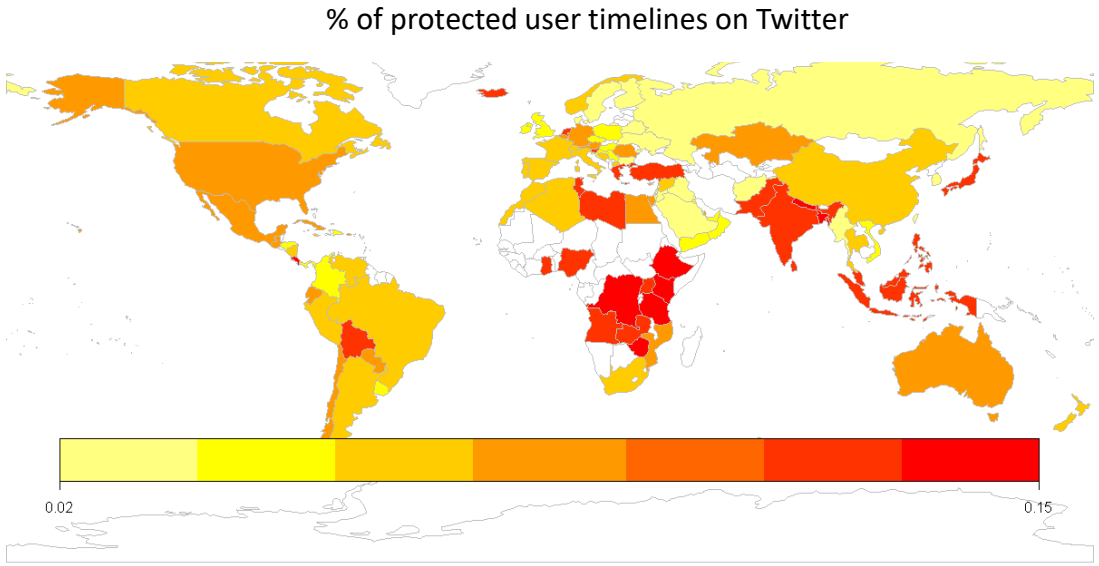
Figure 1. Social media users' activity rhythm more similar to their followees' than to their followers'.

Liang, H. & Shen, F. (2018). Birds of a schedule flock together. *CHB*

1. 使用 – 隐私



Liang, H. & Fu, K. W. (2017). Information overload, similarity, and redundancy. *JCMC*



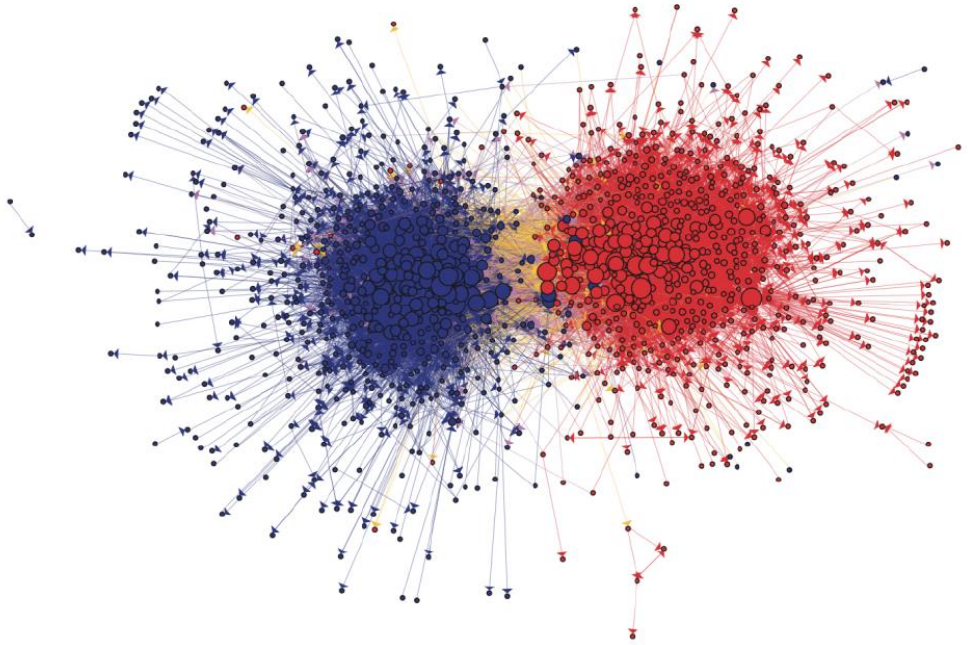
Liang et al., (2017). Privacy protection and self-disclosure across societies. *NMS*

2. 结构 - 网络分析

- 宏观结构
  - 制度、政体、规范、文化等
    →
  - 网络结构：关系网络的结构特征.

- 微观交互（涌现）
  - 自我中心网 → 全网
  - 社会选择与社会影响

2. 结构 - 碎片化



Adamic & Glance (2005). The political blogosphere and the 2004 US election: divided they blog. *ACM*

Table 1 Frequency and Expected Frequency of Cross-Ideological Debate in Political (A) and Nonpolitical (B) Discussions

A From-to-	Centrist	Left	Right	Row Total
Centrist	**9,915** (**9,347**)	12,951 (14,022)	17,906 (17,403)	40,772
Left	12,077 (13,365)	**8,103** (**20,049**)	38,116 (24,882)	58,296
Right	17,288 (16,568)	37,872 (24,855)	**17,110** (**30,847**)	72,270
Column total	39,280	58,926	73,132	171,338
B **From-to-**	**Centrist**	**Left**	**Right**	**Row Total**
Centrist	**19,883** (**19,044**)	8,565 (9,139)	29,075 (29,341)	57,523
Left	12,284 (9,318)	**5,785** (**4,471**)	10,075 (14,355)	28,144
Right	17,807 (21,613)	9,631 (10,371)	**37,884** (**33,298**)	65,282
Column total	49,974	23,981	76,994	150,949

Note: Cell entries are the observed frequencies with expected values in parentheses. Bold values indicate within-ideological discussions. Self-replies are excluded.

Liang, H. (2014). The organizational principles of online political discussion. *HCR*

2. 结构 - 对话规范

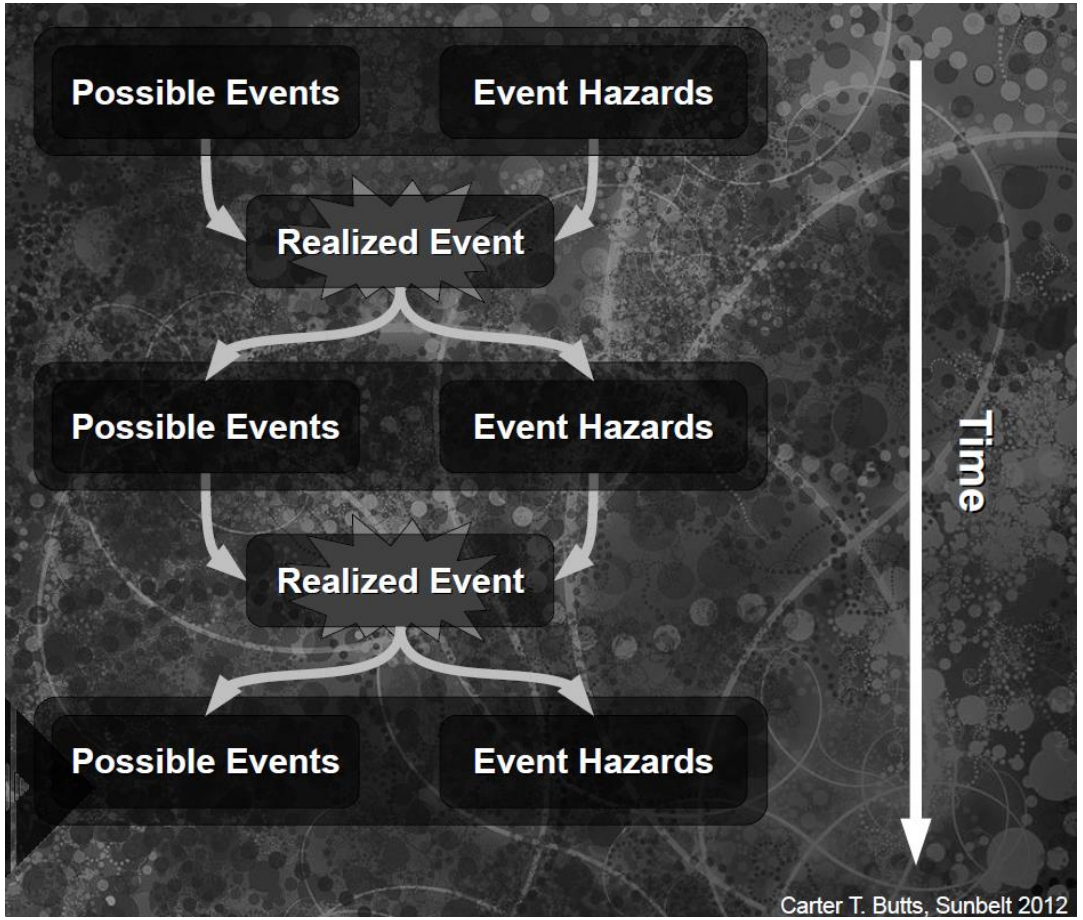


Table 2 Random-Effects Relational Event Model for Discussion Organization

	Model 1	Model 2	Model 3	Model 4
Ideology				
Cross versus within	1.05**	0.76**	0.75**	0.80**
	(.03)	(.02)	(.02)	(.02)
Centrist versus within	0.73**	0.55**	0.53**	0.54**
	(.03)	(.03)	(.03)	(.06)
Conversational norms				
$AB \rightarrow BA$		1.93**	1.89**	1.88**
		(.02)	(.02)	(.02)
$AB \rightarrow BY$		1.13**	1.08**	1.08**
		(.02)	(.02)	(.02)
$AB \rightarrow AY$		1.39**	1.34**	1.34**
		(.02)	(.02)	(.02)
Structure				
Popularity		6.86**	6.61**	0.83**
		(.08)	(.08)	(.08)

Liang, H. (2014). The organizational principles of online political discussion. *HCR*

2. 结构 - 网络与内容冗余性

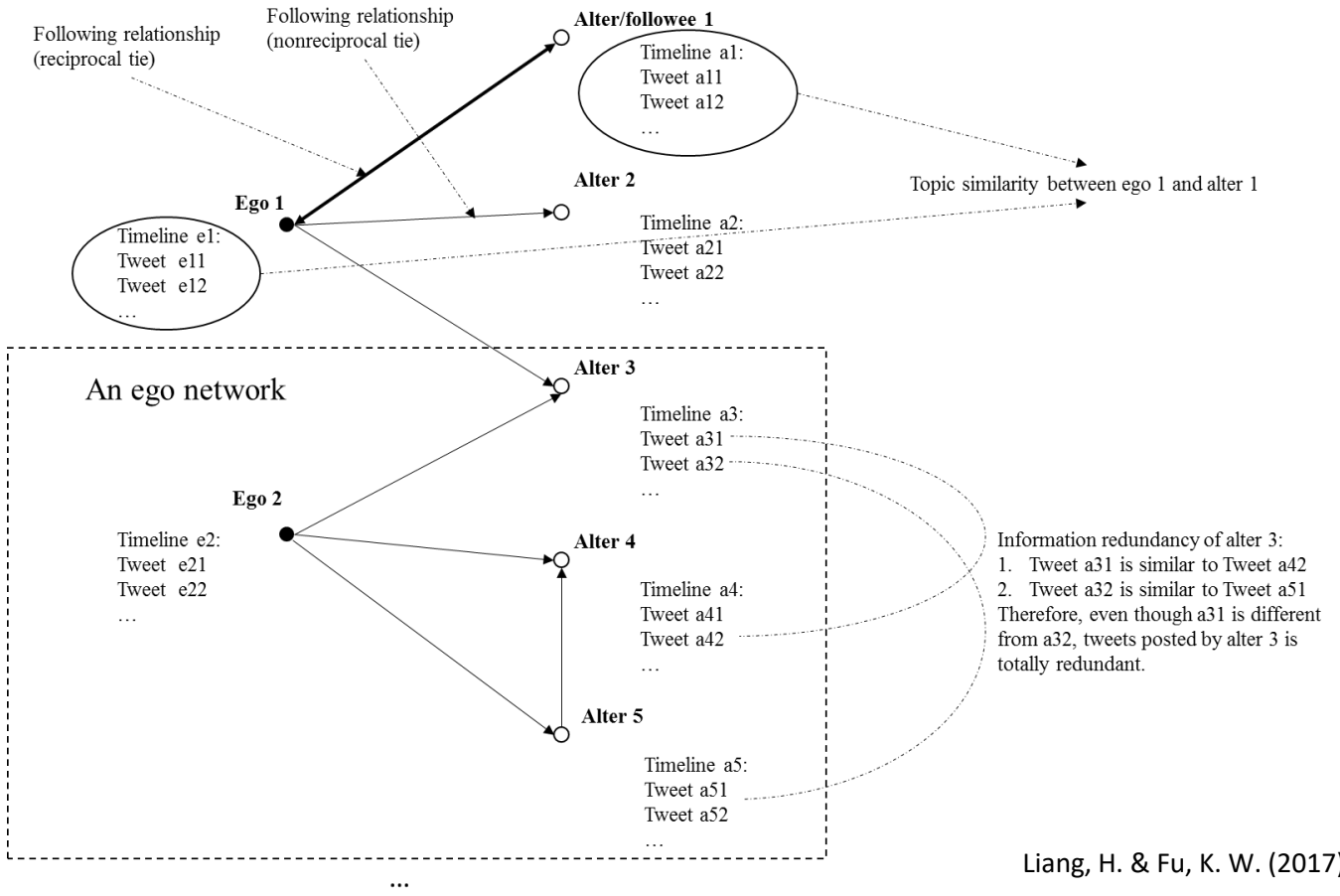
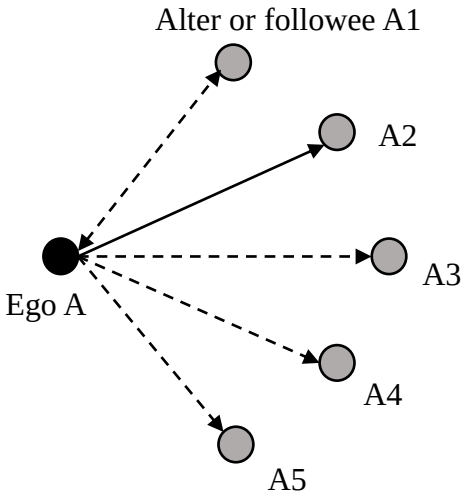


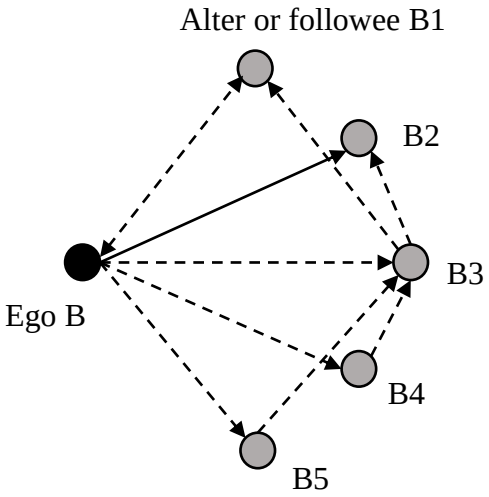
Table 1 Multilevel Logistic Regression Models Predicting Unfollowing

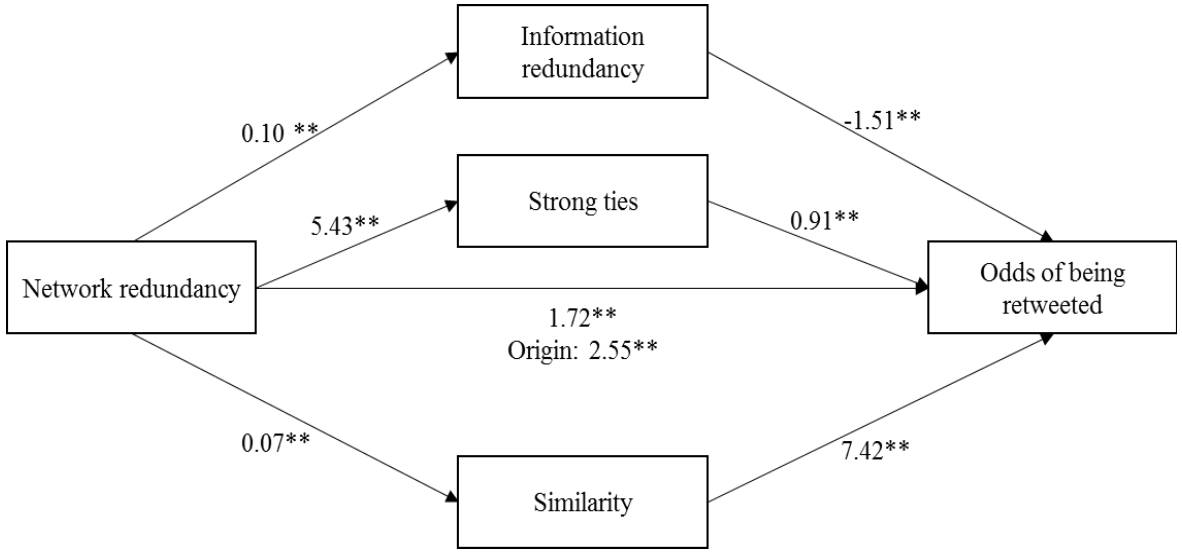
	Model 1		Full Model		Raw Text Measures	
	Estimate (SE)	Z	Estimate (SE)	Z	Estimate (SE)	Z
			Informational Variables			
Similarity	−2.312**(.313)	−7.39	−1.700**(.339)	−5.02	−2.349**(.604)	−3.89
Overload#	0.102**(.023)	11.55	0.083**(.009)	9.18	0.038**(.013)	2.87
Redundancy	−0.019 (.002)	−0.80	0.120**(.024)	5.00	0.071*(.031)	2.26
Overload × Similarity	−0.353*(.141)	−2.50	−0.458**(.139)	−3.28	0.201**(0.77)	2.63
Overload × Redundancy	−0.068**(.018)	−3.82	−0.032 (.018)	−1.79	0.049*(.020)	2.47
Similarity × Redundancy	1.489**(.502)	2.97	0.953 (.562)	1.70	1.868 (1.056)	1.77
			Relational Variables			
Alter popularity#			−0.036**(.009)	−4.14	−0.036**(.009)	−4.19
Reciprocity			−1.518**(.017)	−89.58	−1.513**(.0170)	−89.43
Shared followees#			−0.151**(.007)	−21.23	−0.151**(.007)	−21.16
			Control Variables			
			Model Summary			
Var. of intercepts across users	8.138 (2.853)		9.021 (3.003)		7.177 (2.679)	
Log-Likelihood	−119,115.4		−113,763.1		−113,979.6	
Conditional R^2	67.8%		69.7%		65.2%	
Marginal R^2	2.6%		6.1%		7.0%	
# of ties			1,613,735			
# of users			7,326			

Liang, H. & Fu, K. W. (2017). Information overload, similarity, and redundancy. *JCMC*

2. 结构 – 网络与内容冗余性



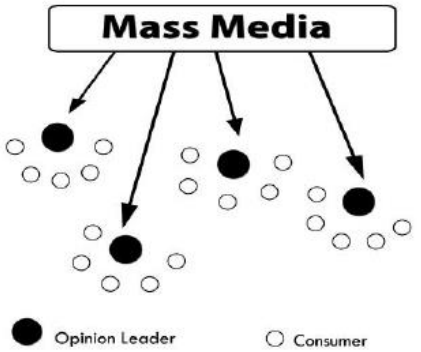




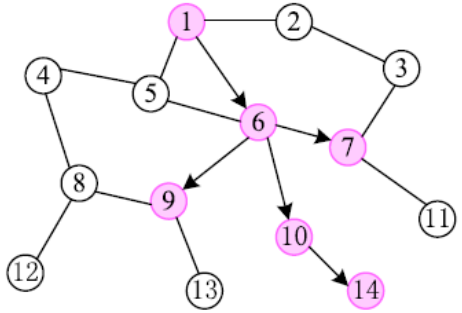
Liang, H. & Fu, K. W. (2019). Network redundancy and information diffusion. *CR*

3. 扩散 – 扩散网络

- 卡茨-拉扎斯菲尔德：两级传播理论



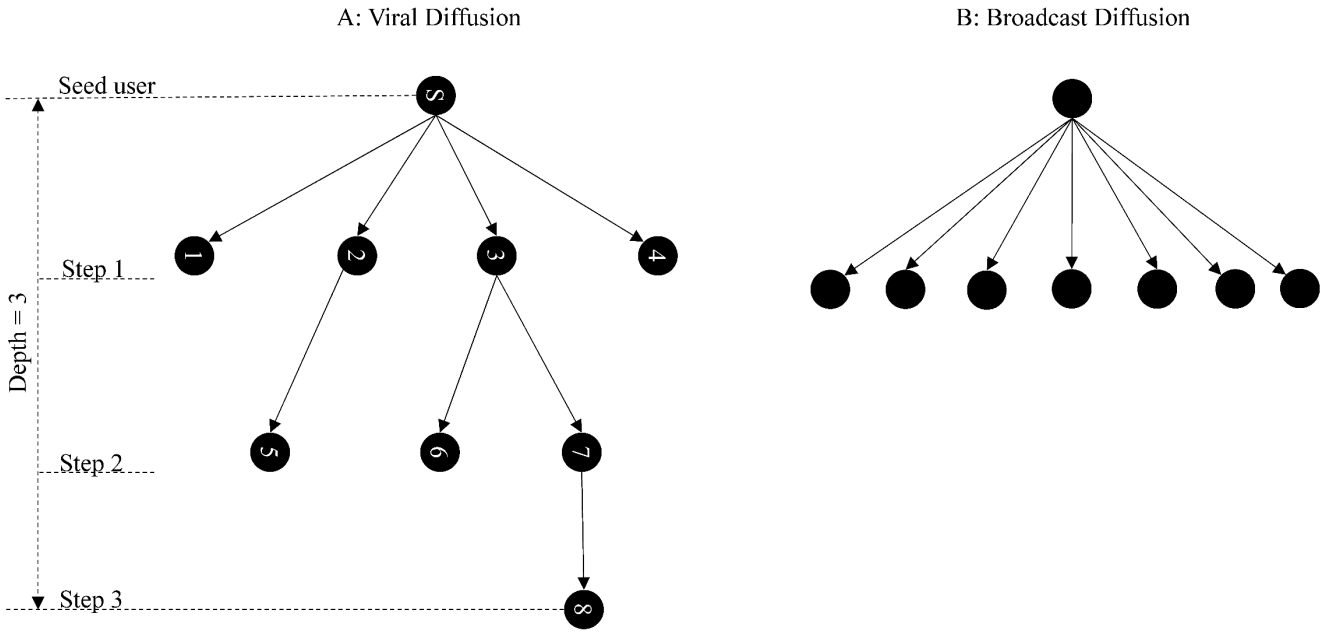
  - 追踪传播路径?
  - 传播网络?
  - 传播路径的形成?



  - 信息扩散 -> 影响力扩散 (personal influence)
  - 社会选择 (homophily-driven diffusion) vs. 社会影响

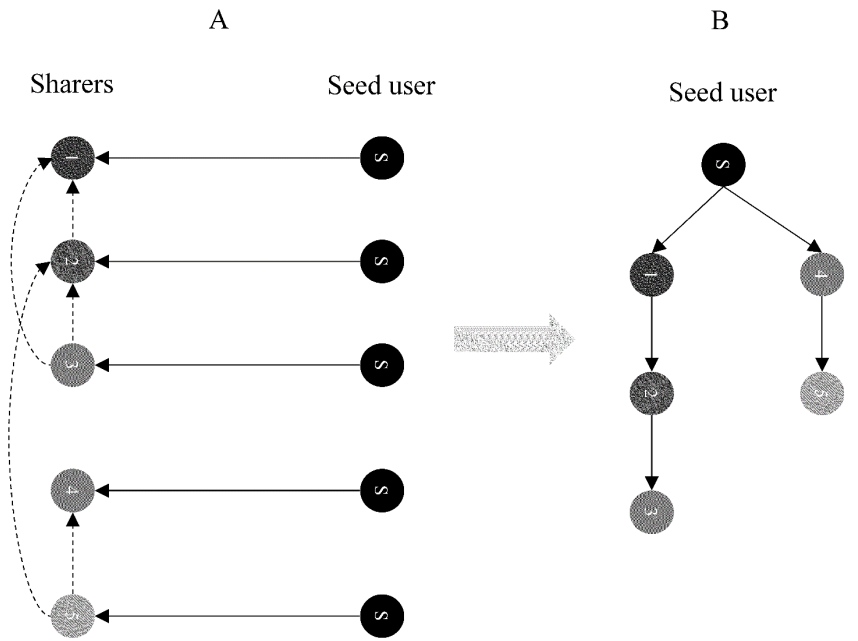
3. 扩散 – 级联

- 传播结构
  - 深度
  - 广度
  - 结构性扩散度 (Goel et al., 2016)
  - …

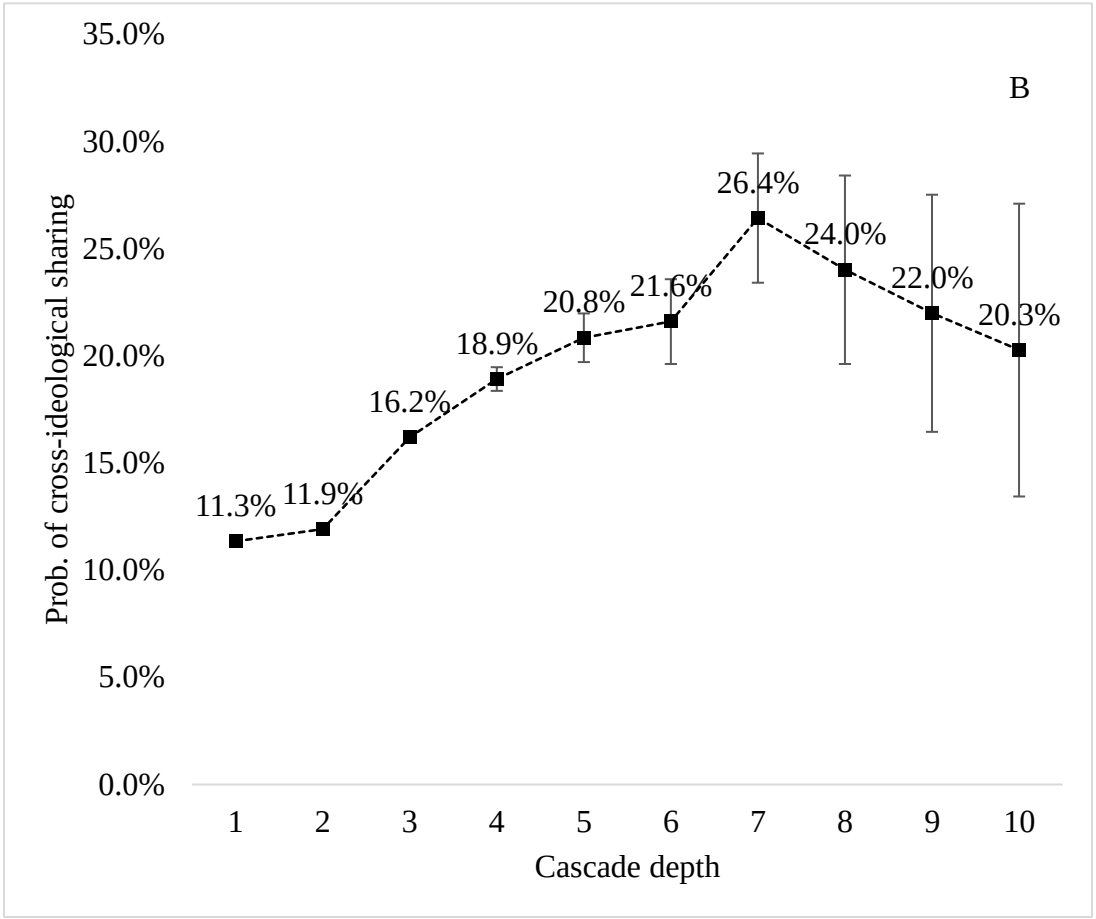


Liang, H. (2018). Broadcast versus viral spreading. *JOC*

3. 扩散 – 级联



Liang, H. (2018). Broadcast versus viral spreading. *JOC*



4. 内容 – 文本挖掘

- 作为测量的文本挖掘
  - 规模化传统方法
  - 测量已有概念
  - 创造新的概念

- 作为分析的文本挖掘
  - 文本作为因变量/自变量/混淆变量
  - 与其它方法相结合：文本+网络分析

4. 内容 – 共同话语基础

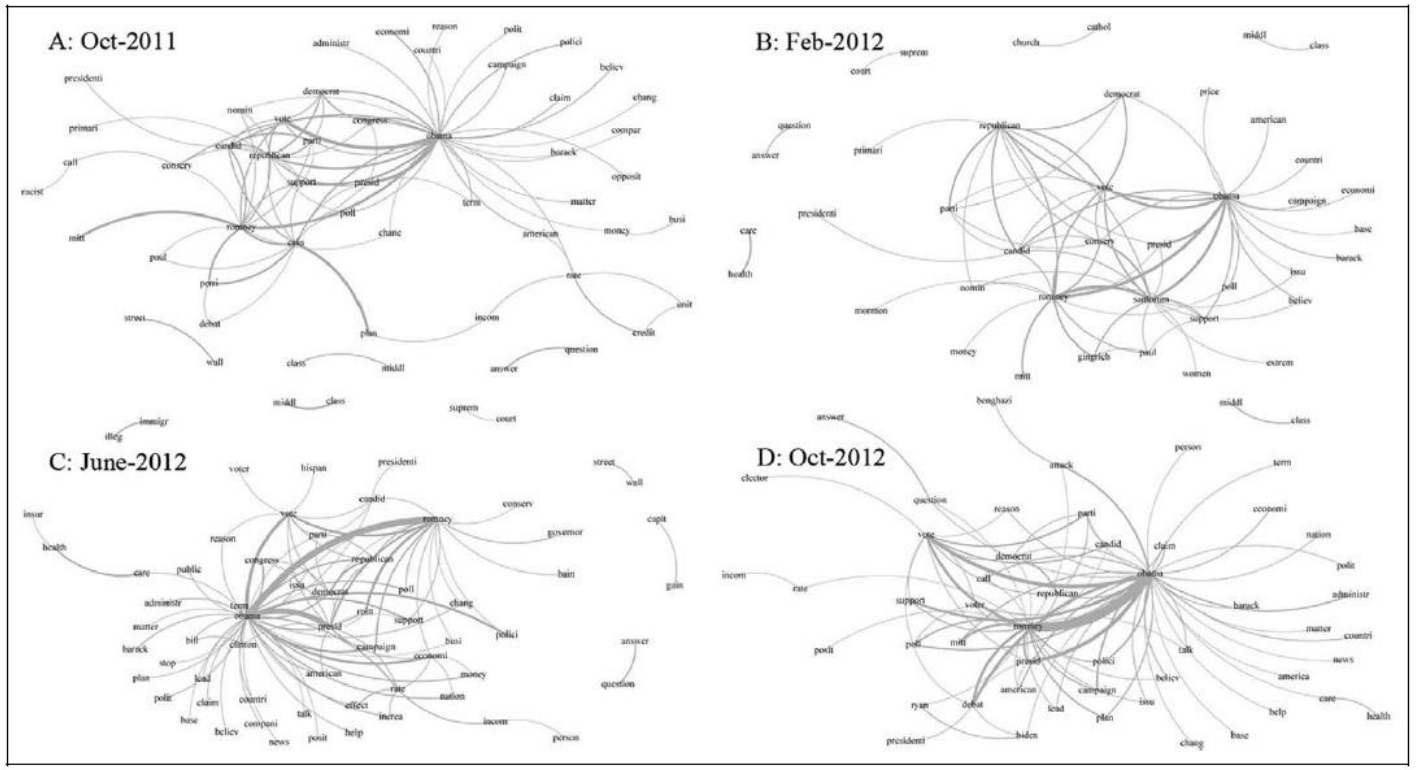
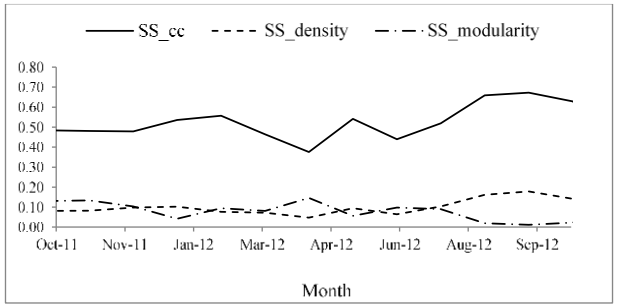
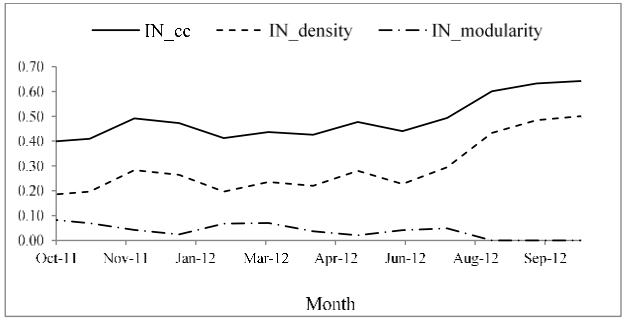


Figure 2. The interpretive networks over time. October 2011 (modularity is 0.09), February 2012 (modularity is 0.07), June 2012 (modularity is 0.04), and October 2012 (modularity is 0.00).



B. The evolution of semantic network



C. The evolution of interpretive network

Figure 2. The evolution of discussion and cognitive networks over 13 months.

Liang, H. (2014). Coevolution of political discussion and common ground in web discussion forum. *SSCR*

4. 内容 – 网络 × 共同话语基础

- 协同演化模型 (bipartite network)
  - *H1* is related to whether agreement leads to discussion ($\{Y_{ia} + Y_{ja}\} \Rightarrow X_{ij}$)
  - *H2* focuses on whether discussion leads to agreement ($\{X_{ij} + Y_{ia}\} \Rightarrow Y_{ja}$)

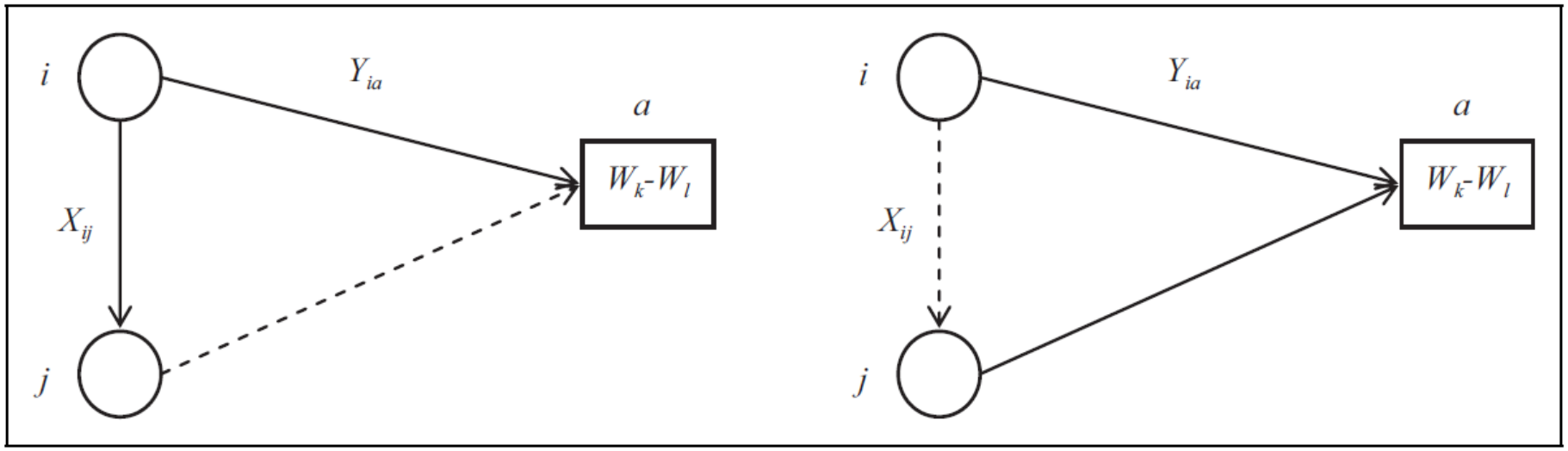


Figure 1. The coevolution of discussion and interpretive network.

Liang, H. (2014). Coevolution of political discussion and common ground in web discussion forum. *SSCR*

4. 内容 – 网络 × 共同话语基础

Table 2. Coevolution of Discussion–Interpretive Network.

Effects	T1 = >T2		T2 = >T3		T3 = >T4	
	Parameter	*SE*	Parameter	*SE*	Parameter	*SE*
Discussion network						
Rate	64.52**	11.40	61.57**	4.74	62.87**	11.08
Density	–5.77**	0.21	–6.25**	0.36	–5.37**	0.39
Reciprocity	3.53**	0.26	4.16**	0.36	3.17**	0.26
Transitivity	–0.03	0.02	–0.12**	0.02	–0.05*	0.02
Popularity	–0.13**	0.02	–0.15**	0.01	–0.06**	0.01
Same ideology	–0.21**	0.08	–0.31**	0.09	–0.42**	0.12
Activity alter	0.21**	0.02	0.25**	0.02	0.17**	0.01
Activity ego	0.10**	0.01	0.12**	0.01	0.13**	0.03
Same frame	0.01**	0.00	0.01**	0.00	0.01*	0.00
Interpretive network						
Rate	66.27**	5.08	45.24**	3.43	50.28**	7.01
Density	–2.46**	0.23	–2.76**	0.22	–2.70**	0.41
Popularity	–0.00	0.00	–0.00	0.00	0.00	0.00
Out-in assortativity	0.03**	0.00	0.03**	0.00	0.03**	0.00
Out-degree activity	–0.26*	0.12	–0.59**	0.17	–0.42*	0.19
Discussion	0.11*	0.05	0.21**	0.06	0.16*	0.07

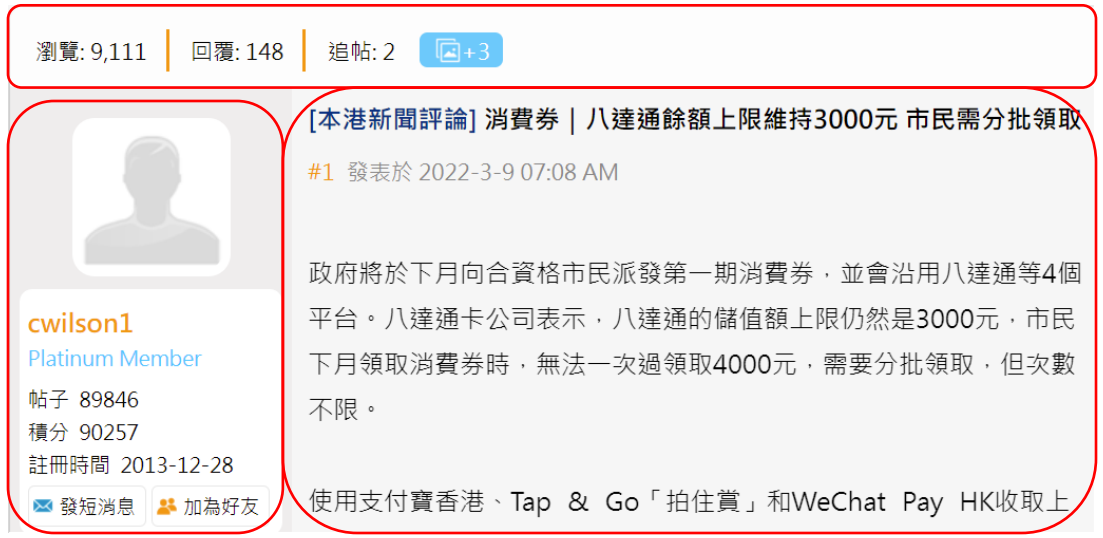
Liang, H. (2014). Coevolution of political discussion and common ground in web discussion forum. *SSCR*

5. 效果 – .*

- 观察性数据
  - 自然实验
  - 因果推断 (counterfactual: RCM/PCM)
  - (digital) 现场试验

- 文本（内容）效果
  - 自变量 (e.g., text effect)
  - 混淆变量 (e.g., high dimensional control through deep learning)
  - 因变量 (e.g., predicting topic prevalence)

5. 效果 – 不文明言语之后果

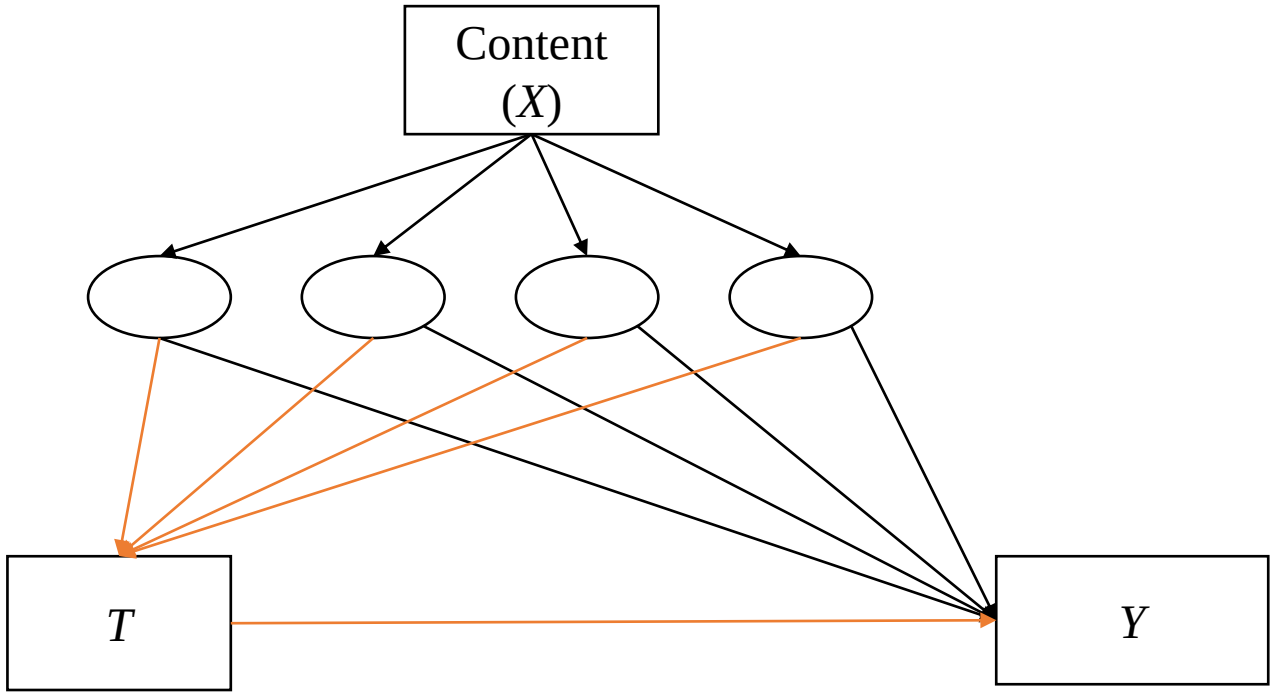


- Incivility → Engagement/参与 (#/%)
- 用户
  - Influence, ability, etc.
  - Repeated observations → Fixed effect
- 背景
  - Timing: competing threads, audience availability etc.
- 文本
  - Other than “uncivil” words
  - High dimensional
  - Double machine learning (Chernozhukov et al., 2018)
  - Doubly robust learner (Kennedy, 2020)

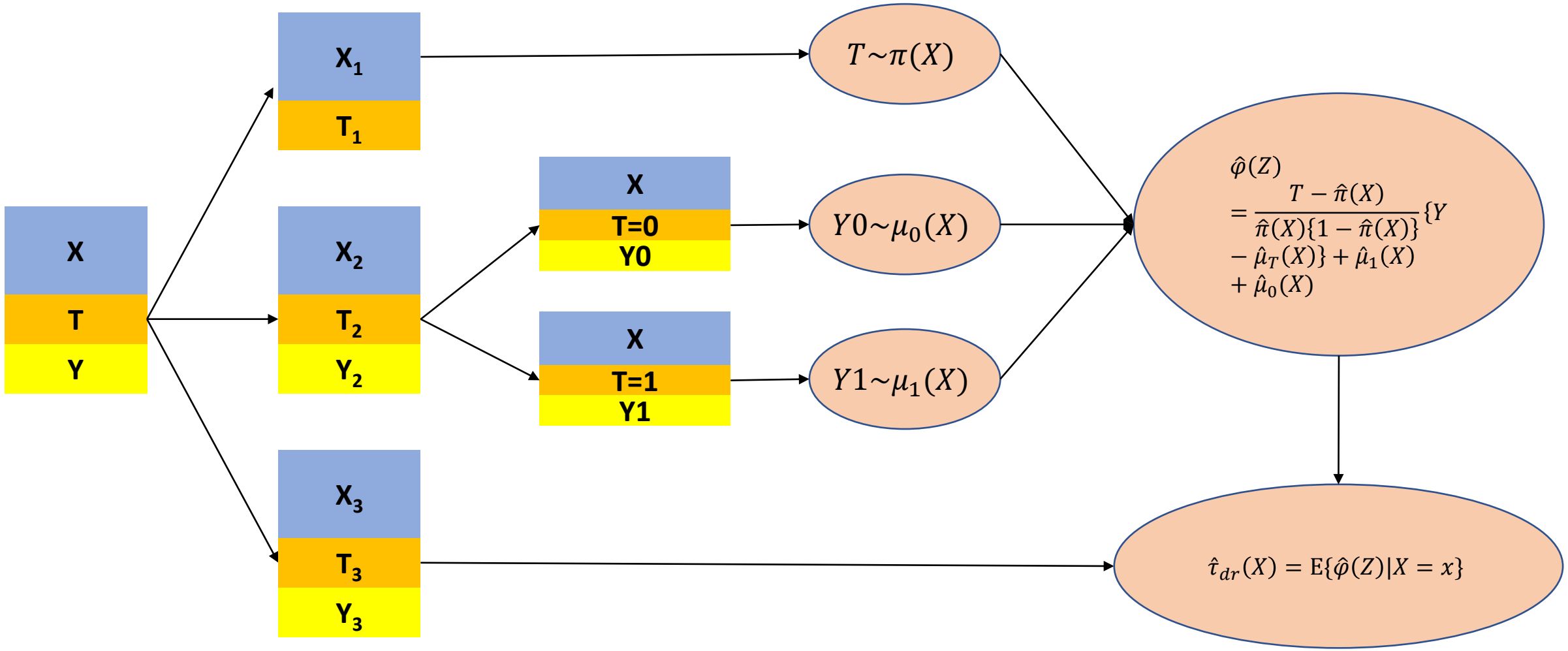
5. 效果 – Double Machine Learning

- Double Machine Learning

1. $M_y(Y \sim X)$
2. $M_t(T \sim X)$
3. $\hat{Y} = Y - M_y(X); \hat{T} = T - M_t(X)$
4. $\hat{Y} = \alpha + \tau\hat{T} + fe(id) + ?$

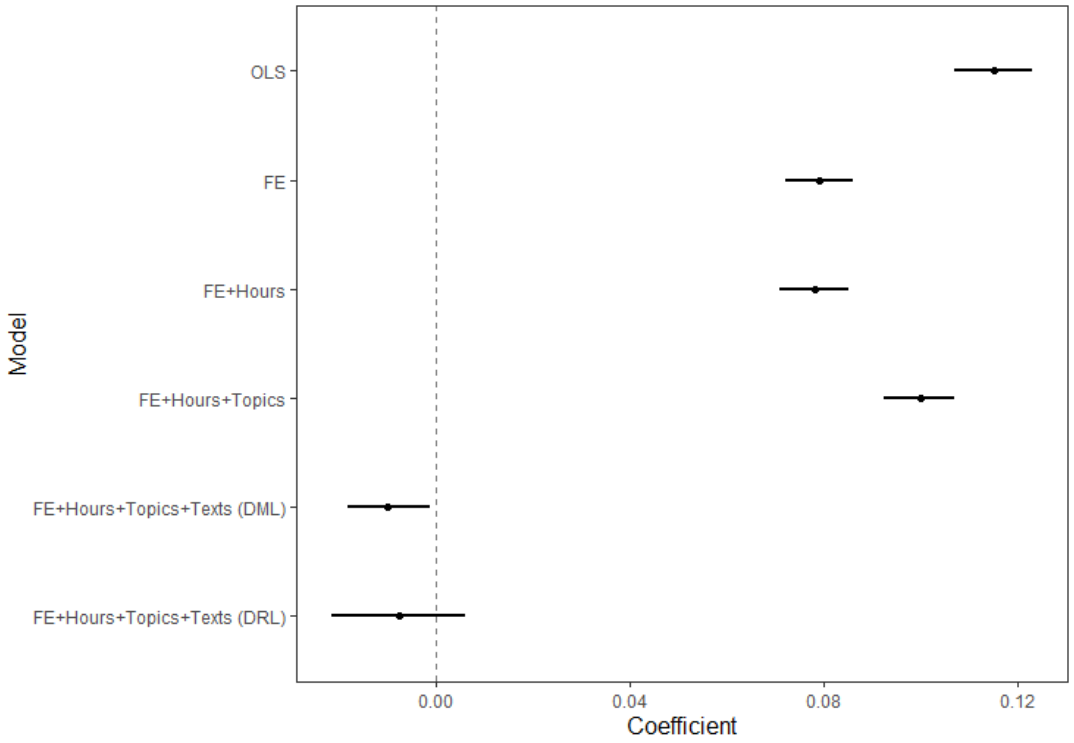


5. 效果 – Doubly Robust Learner

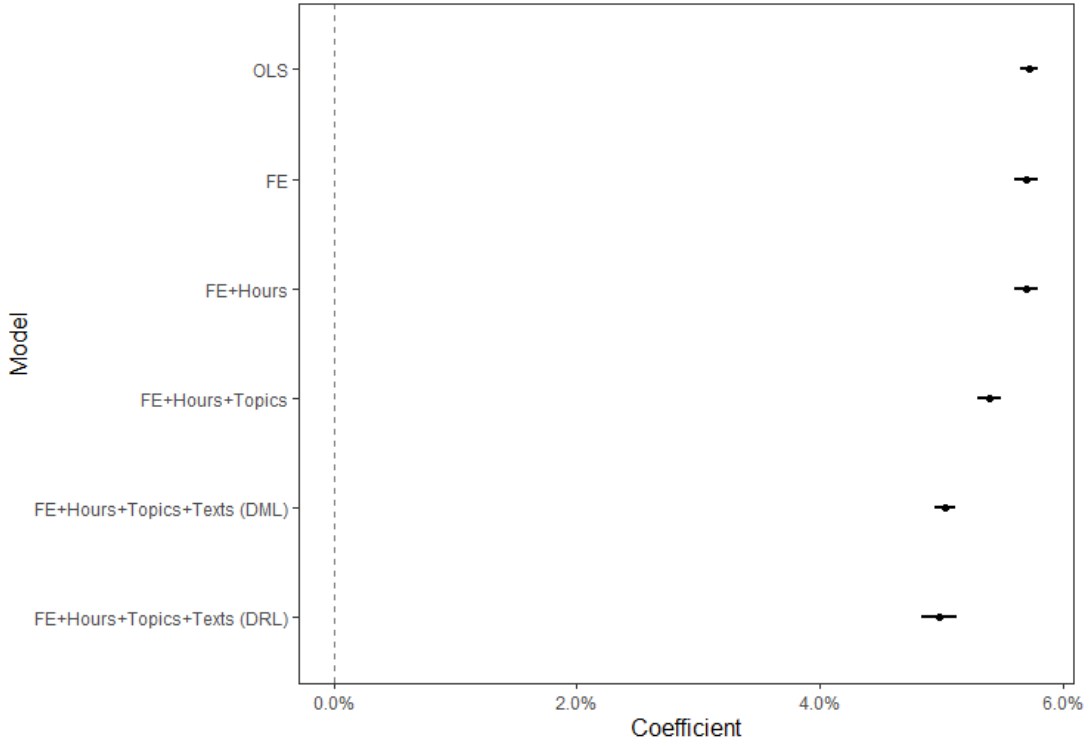


5. 效果 – 不文明言语之后果

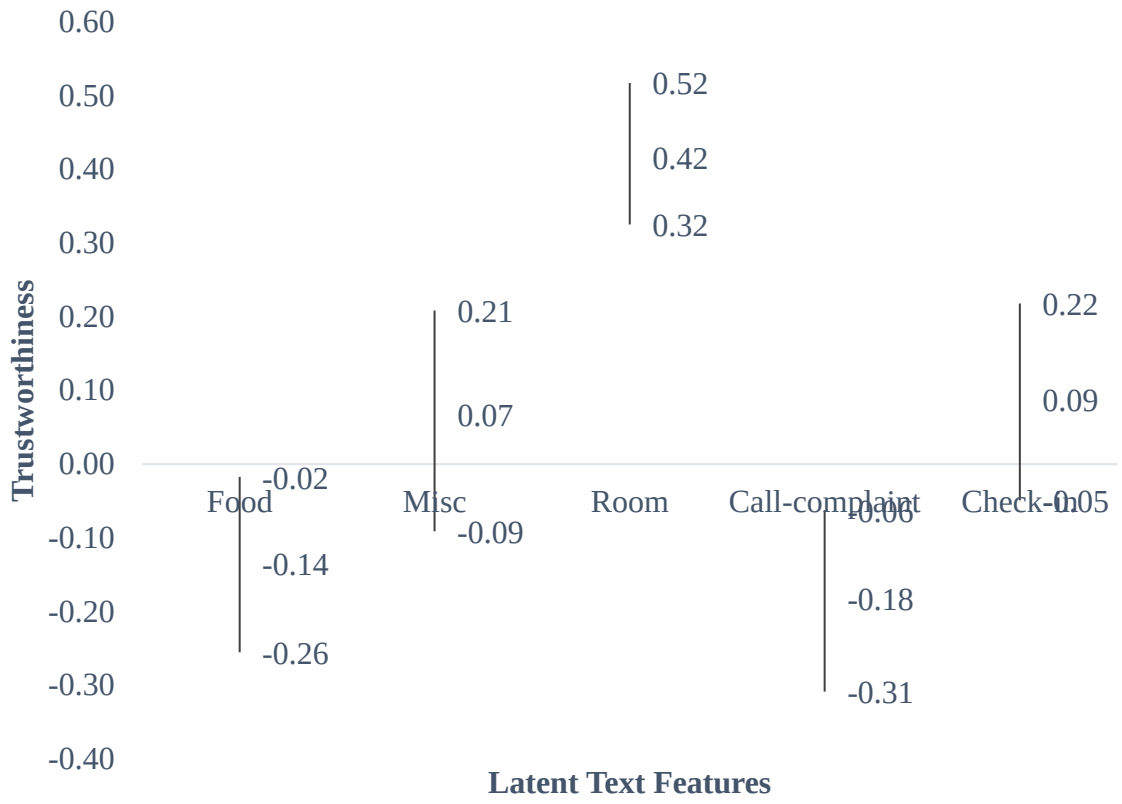
log(#replies) (engagement) LIHKG



%uncivil_replies (contagion) LIHKG



5. 效果 – (Latent) 文本效果



No.	Keywords	Manual label
1	bread, pool, house, resort, favorite, drive, sausage, chicken, politely, think	Food
2	inform, fact, double, fund, yell, reach, remainder, final, instead, deposit	Misc.
3	room, time, floor, one, bathroom, two, area, can, though, use	Room
4	call, tell, manager, speak, say, check, ask, situation, finally, call_back	Call-complaint
5	one, hotel, room, make, offer, see, think, get, check-in, include	Check-in

Huang, G. & Liang, H. (2021). Textual effects of consumer reviews. *JBR*

一些思考

- 计算传播学有什么理论贡献？
  - 计算社会科学在现代化（modernize），而非取代传统社会科学。
  - 计算方法是社会科学研究方法中的一种。
  - 新的理论往往伴随新的方法（研究设计），但这并非必然。
  - 还需不需要理论？
  - 是技术/现象还是方法带来/需要新的理论？


NEW 新书推荐
大数据时代的计算舆论学：
理论、方法与案例
周葆华 梁 海 主编
扫码进入天猫商店
复旦大学出版社
Fudan University Press