Social CyberSecurity

徐冰冰

目录

Social Cybersecurity

社交媒体众多

社交机器人、虚假身份、
真实用户并存

极端观点　　自由言论

政治舆论　　...

- Beliefs opinions and attitudes are shaped
- Indistinguishable between true and false

Two Objectives

Social cybersecurity is an applied computational social science with two objectives

- **characterize, understand, and forecast** cyber-mediated changes in human behavior and in social, cultural, and political outcomes

- build a social cyber infrastructure that will **allow the essential character of a society to persist** in a cyber-mediated information environment that is characterized by changing conditions, actual or imminent social cyberthreats, and cyber-mediated threats

相似学科

Cybersecurity

- **focused on machines**, and how computers and databases can be compromised.
- understand the technology, computer science, and engineering

Social cybersecurity

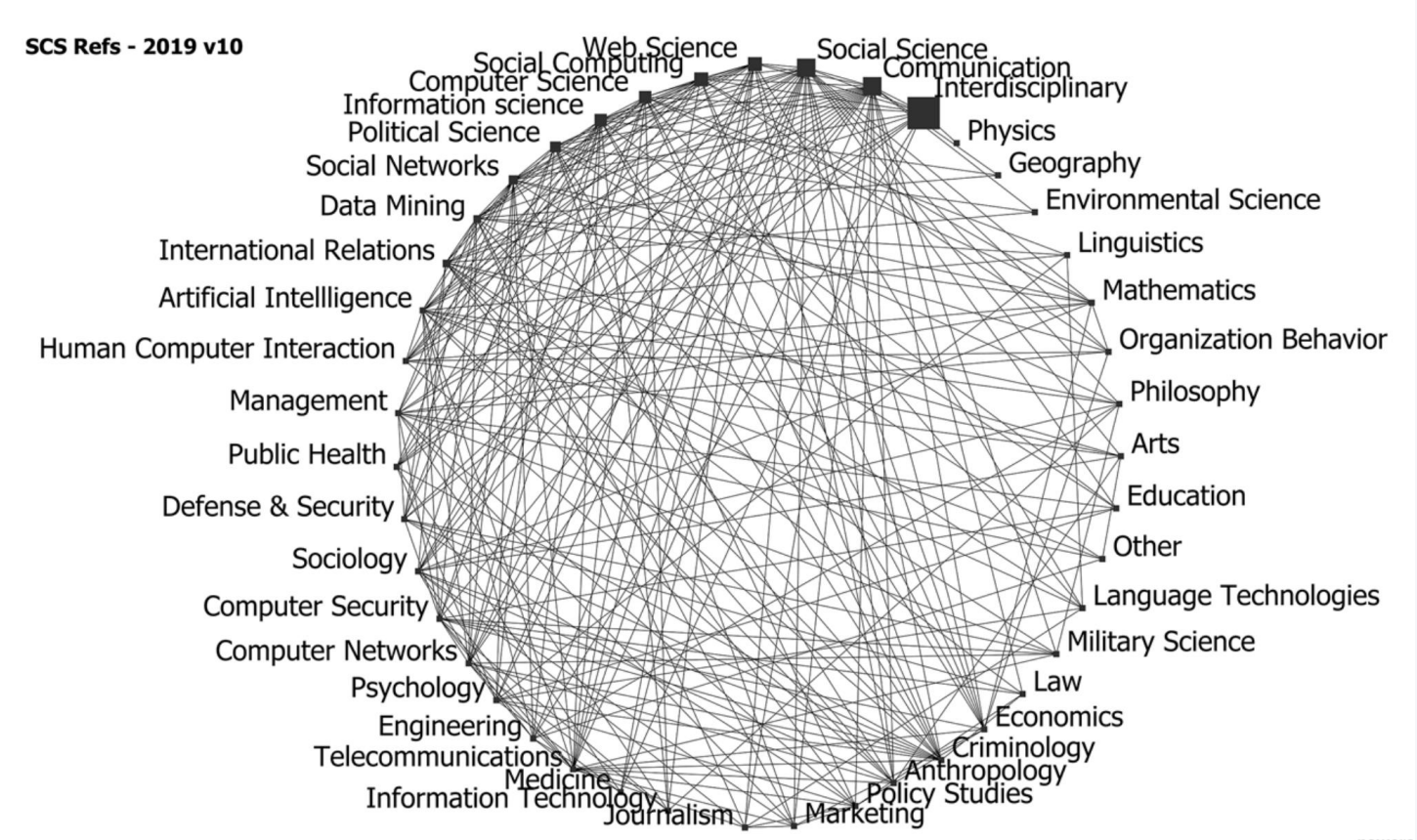
- **focused on humans** and how these humans can be compromised, converted, and relegated to the unimportant, how the digital environment can be manipulated to alter both the community and the narrative
- understand social communication and community building, statistics, social networks, and machine learning

Cognitive security

- **focused on human cognition** and how messages can be crafted to take advantage of normal cognitive limitations.
- understand psychology

学科交叉

- **build on work in high dimensional network analysis, data science, machine learning, natural language processing and agent-based simulation.**



Popular themes

The popular themes in social cybersecurity

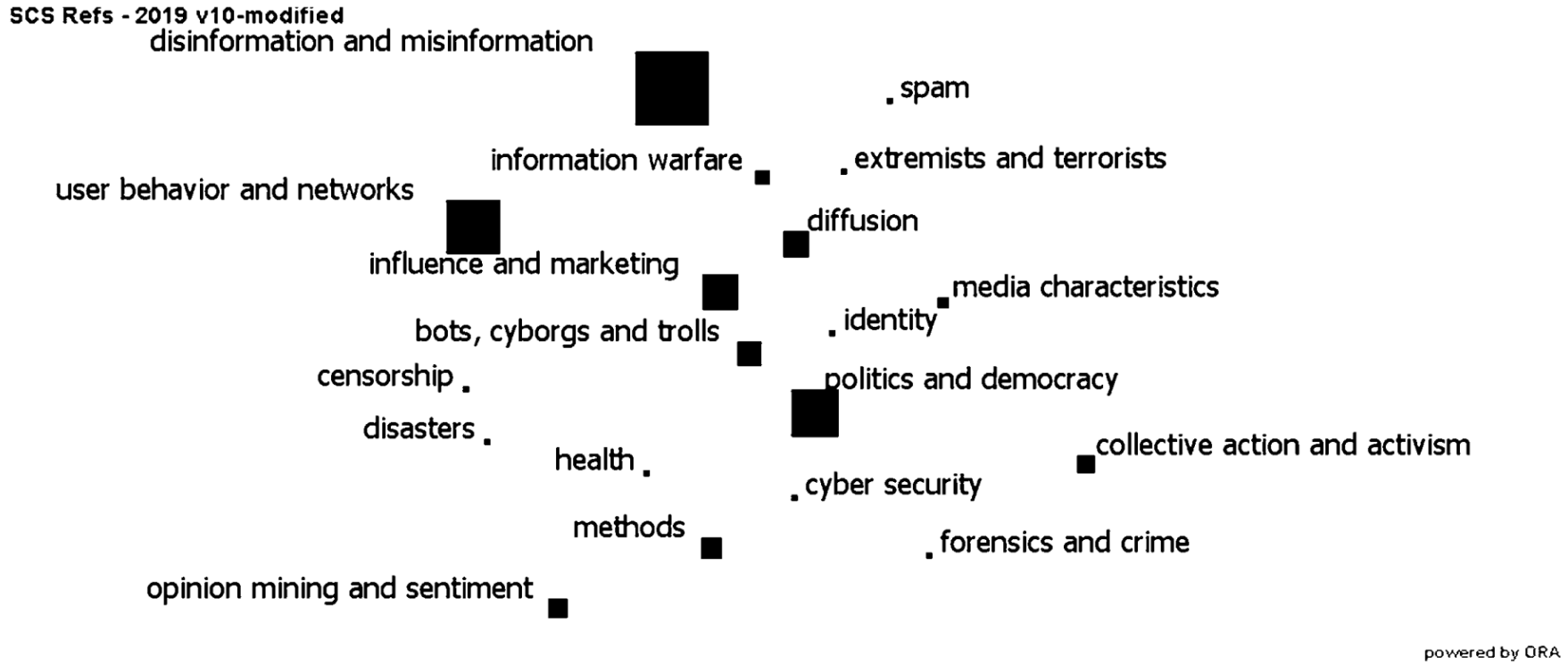


Fig. 2 Research topic areas in social cybersecurity

Case study1: building community in social media(Ukraine)

- **A group of young men sending out provocative images of women. They didn’t know each other they were just posting images they liked**
- **Bots were used in an infuence campaign to send out tweets mentioning each other and multiple of these young men at once**
- **This led the men to learn of others who, like them, were sending out these images. They formed an online group—a topic ori_x0002_ented community. Once formed, the bots now-and-then tweeted information about where to get guns, ammunition, and how to get involved in the fght.**
- **topic oriented communities—groups of actors all communicating with each other about a topic of interest**
- **At the same time, Bots conducted an “enhance” campaign by rebroadcasting some images and pointing to others and an “excite” campaign with new positive language.**

Bots Based **communities**

整体脉络

- **Provide evidence about who is manipulating social media and the internet for/against you or your organization, what methods are being used, and how these social manipulation methods can be countered**.

Table 1 Communication objectives BEND

	Manipulating the narrative		Manipulating the social network	
Positive	**Engage**	Messages that bring up a related but relevant topic	**Back**	Actions that increase the importance of the opinion leader or create a new opinion leader
	Explain	Messages that provides details on or elaborate the topic	**Build**	Actions that create a group or the appearance of a group
	Excite	messages that elicit a positive emotion such as joy or excitement	**Bridge**	Actions that build a connection between two or more groups
	Enhance	Messages that encourage the topic-group to continue with the topic	**Boost**	Actions that grow the size of the group or make it appear that it has grown
Negative	**Dismiss**	Messages about why the topic is not important	**Neutralize**	Actions decrease the importance of the opinion leader
	Distort	Messages that alter the main message of the topic	**Nuke**	Actions that lead to a group being dismantled or breaking up, or appearing to be broken up
	Dismay	Messages that elicit a negative emotion such as sadness or anger	**Narrow**	Actions that lead to a group becoming sequestered from other groups or marginalized
	Distract	Discussion about a totally different topic and irrelevant	**Neglect**	Actions that reduce the size of the group or make it appear that the group has grown smaller

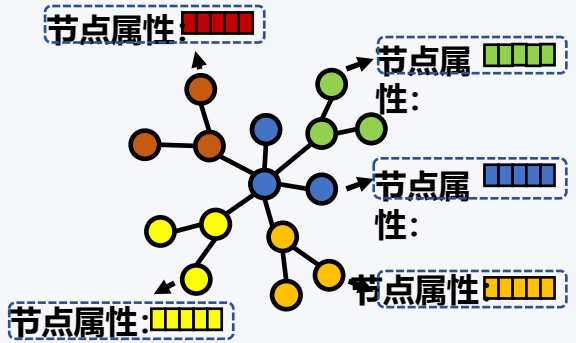
目录

- Social Cyber security：an emerging science
- 复杂网络的研究路线
  - 复杂网络的结构功能性质及其应用（汪小帆）
  - 观测、建模、预测、控制、应用
- 观点动力学预测
  - Predicting Opinion Dynamics via Sociologically-Informed Neural Networks（KDD2022）

整体脉络

社交网络

个体：Identity → 节点：Node
群体：Group → 图：Graph
影响：Influence → 结构：Edge / Structure
外显因素：Portraint → 属性：Attribute
内在功能：观点 → 标签：Label / Function



属性图
(Attribute Graph)

观测：复杂网络的基本结构性质及其度量方法
重点：揭示刻画网络系统结构的基本性质，以及度量这些性质的合适方法.

建模：复杂网络结构的产生机理及其建模
重点：建立合适的网络模型以理解网络结构的产生机理以及网络拓扑性质的意义，并用于预测和控制网络行为

预测：复杂网络的结构与功能之间的关系
重点：基于单个节点的特性和整个网络的结构性质分析与预测网络的行为

控制：改善网络功能的有效方法
重点：提出改善已有网络性能和设计新的网络的有效方法

应用：复杂网络理论的典型应用
重点：关键基础设施网络、生物网络、经济与社会网络

整体脉络

- 复杂网络的基本结构性质及其度量方法 （观测）
  重点：揭示刻画网络系统结构的基本性质，以及度量这些性质的合适方法.
  哪些拓扑性质对于刻画网络结构具有基本的重要性？（拓扑性质决定网络结构？某拓扑性质下相似的两个网络其结构相似，或者其他拓扑性质也相似？）
  大规模网络的结构性质的有效的度量方法（如何高效计算拓扑性质）
  对于随时间演化的网络的结构性质分析

- 复杂网络结构的产生机理及其建模 （建模）
  重点：建立合适的网络模型以理解网络结构的产生机理以及网络拓扑性质的意义，并用于预测和控制网络行为
  建模的两个极端：再现模型（假设多，参数多，难解释，难预测）和概念模型（不考虑实际网络，难应用），如何基于对实际网络的理解，找到上述的均衡？
  图论+统计物理（现阶段是否有数据驱动？）
  传统的图论是否仍然适于处理这类演化网络？是否有可能建立更为合适的新的描述与分析工具？

整体脉络

- 复杂网络的结构与功能之间的关系（预测）
  重点：基于单个节点的特性和整个网络的结构性质分析与预测网络的行为，（了解网络结构对网络功能的影响）结构是由多种拓扑性质共同决定的，可以判断网络某个特定功能和某个拓扑性质之间的关系，但难以准确判断该功能是否是由网络的某个结构性质所决定的（需要假设网络所有其它的性质都保持不变的情况下，考虑该性质的变化对网络功能的影响）

- 改善网络功能的有效方法（控制）
  - 提出改善已有网络性能和设计新的网络的有效方法
  - 利用控制行为使复杂网络具有期望功能（通过对复杂网络的少部分节点施加控制从而使整个网络具有期望功能，能控制的节点集合有限情况下？控制手段有限？节点排序？）

- 复杂网络理论的典型应用（应用）
  重点：关键基础设施网络、生物网络、经济与社会网络

目录

- Social Cybersecurity：an emerging science

- 复杂网络的研究路线
  - 复杂网络的结构功能性质及其应用（汪小帆）
  - 观测、建模、预测、控制、应用

- 观点动力学预测
  - Predicting Opinion Dynamics via Sociologically-Informed Neural Networks（KDD2022）

Background

- **Opinion dynamics is the study of information and evolution of opinions in human society. In society, people exchange opinions on various subjects including political issues, new products, and social events (e.g., sports events). As a result of such interactions (i.e., social interaction), an individual’ s opinion is likely to change over time.**

- **Understanding the mechanisms of social interaction and predicting the evolution of opinions are essential**

Related Work

Opinion dynamics model

- Methods
  - DeGroot
  - French
  - bounded confidence models
    - the tendency of people to accept only information that confirms prior beliefs.
- The theoretical models are highly interpretable and can utilize the wealth of knowledge from social sciences

Data-driven methods have been applied to exploit large_x0002_scale data from social media for predicting the evolution of users' opinions.

- it is largely agnostic to prior scientific knowledge of the social interaction mechanism

Motivation

Integrates both large-scale data and prior scientific knowledg

- **first reformulate opinion dynamics models into ordinary differential equations (ODEs). We then approximate the evolution of individuals' opinions by a neural network.**
- **During the raining process, by penalizing the loss function with the residual of ODEs that represent the theoretical models of opinion dynamics, the neural network approximation is made to consider sociological and social psychological knowledge**

Motivation

Challenges

- **they cannot utilize additional side information including individuals' profiles, social connections.**
  - **we combine the framework of PINNs and natural language processing techniques by adding a pre-trained language model**

- **they cannot consider structural knowledge on social interaction.**
  - we apply low-rank matrix factorization to the parameters of ODEs.

Problem Definition

- a social media post is represented as the triple (u, t, y), which means user u made a post with opinion y, on a given subject matter, at time t.
- We denote H = $\{(ui, ti, yi)\}i$=1 as the sequence of all posts made by the user up to time T .A collection of user profiles by D = $\{d1, ..., dU\}$.
- Given the sequence of opinions H during a given time-window [0,T) and user profiles D, we aim to predict users' opinions at an arbitrary time $t*$ in the future time window [T ,T + ΔT].

Method

DeGroot model.

$$x_u(t+1) = x_u(t) + \sum_{v \in \mathcal{U} \setminus u} a_{uv} x_v(t),$$

The DeGroot model captures the concepts of assimilation — the tendency of individuals to move their opinions towards others.

DeGroot model. For the DeGroot model [10], we can transform the difference equation of Equation (1) into an ODE as follows:

$$\frac{dx_u(t)}{dt} = \sum_{v \in \mathcal{U} \setminus u} a_{uv} x_v(t), \tag{5}$$

where t is time and $x_u(t)$ is the user u's latent opinion at time t.

Model Formulation

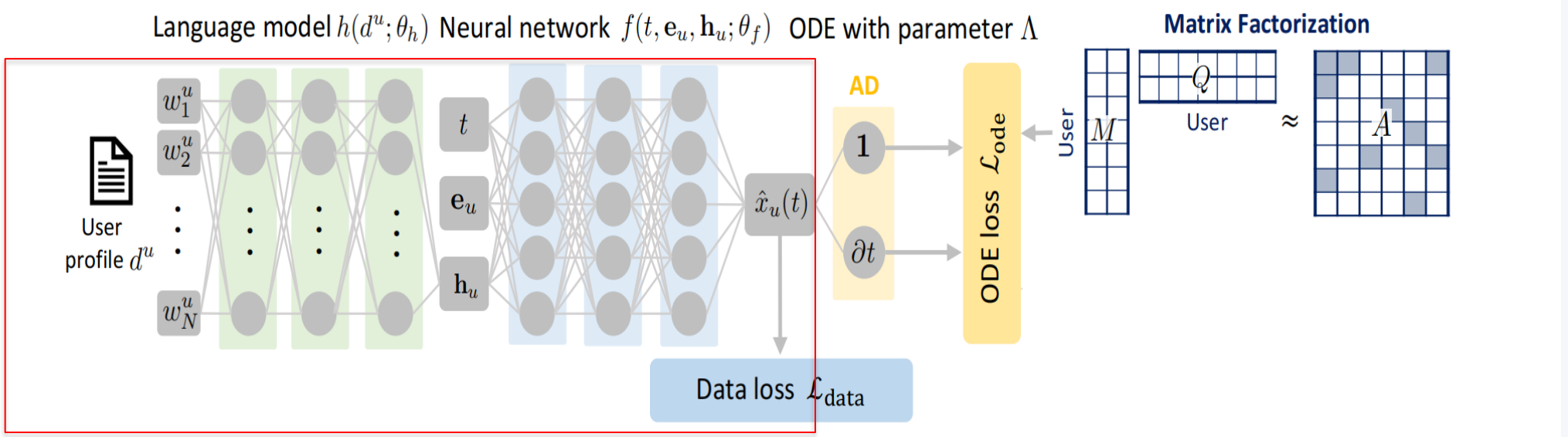


Figure 1: Overall architecture of our proposed method, SINN (Sociologically-Informed Neural Network).

$$\mathcal{L}(\theta_f, \theta_h, \Lambda; \mathcal{H}, \mathcal{D}) = \mathcal{L}_{\text{data}}(\theta_f, \theta_h; \mathcal{H}, \mathcal{D}) + \alpha \mathcal{L}_{\text{ode}}(\theta_f, \theta_h, \Lambda) + \beta \mathcal{R}(\Lambda), \tag{11}$$

Problem Definition

$$\mathcal{L}(\theta_f, \theta_h, \Lambda; \mathcal{H}, \mathcal{D}) = \mathcal{L}_{\text{data}}(\theta_f, \theta_h; \mathcal{H}, \mathcal{D}) + \alpha \mathcal{L}_{\text{ode}}(\theta_f, \theta_h, \Lambda) + \beta \mathcal{R}(\Lambda), \tag{11}$$

$$\mathcal{L}_{\text{data}}(\theta_f, \theta_h; \mathcal{H}, \mathcal{D}) = \frac{1}{I} \sum_{i=1}^{I} L_{\text{CE}}(\hat{y}_i, y_i), \tag{12}$$

$$\mathcal{L}_{\text{ode}}(\theta_f, \theta_h, \Lambda) = \frac{1}{J} \sum_{j=1}^{J} \sum_{u \in \mathcal{U}} \left(\frac{d\hat{x}_u(t)}{dt}\bigg|_{t=\tau_j} - \sum_{v \in \mathcal{U} \setminus u} \mathbf{w}_u^\top \mathbf{h}_v \hat{x}_v(\tau_j) \right)^2, \tag{13}$$

$$\mathcal{R}(\Lambda) = \|M\|_1 + \|Q\|_1$$

During training, enforce the FNN output $\hat{x}_u(t)$ to

- reproduce the ob_x0002_served opinions
- satisfy the governing ODEs that represent the theoretical models of opinion dynamics.

Experiments

Table 3: F1 score (F1) and Accuracy (ACC) for predicting opinions from three real-world datasets. Higher is better. The best performance is highlighted in bold.

	Twitter BLM		Twitter Abortion		Reddit Politics	
	ACC	F1	ACC	F1	ACC	F1
Voter	0.199	0.163	0.222	0.170	0.628	0.500
DeGroot	0.203	0.131	0.358	0.203	0.807	0.389
AsLM	0.092	0.117	0.435	0.195	0.789	0.441
SLANT	0.105	0.070	0.425	0.175	0.733	0.496
SLANT+	0.091	0.042	0.437	0.152	0.789	0.441
NN	0.336	0.237	0.441	0.369	0.875	0.824
Proposed	**0.359**	**0.246**	**0.467**	**0.412**	**0.927**	**0.884**

Discussion

论文还是采用了比较传统的PINN的方式，也就是用loss的方式建模ODE约束，ODE约束了两个方面，

- 观点的时序依赖，但在前向预测每个时刻的观点时，没有考虑时刻之间的关联（即没有用t_0到t_i去预测t_{i+1}）
- 观点和邻居节点的关系也是通过约束实现（可以用GNN建模，而不需要通过loss约束）

思路：用求积分和GNN的方式去设计神经网络，不只把ODE作为loss，而是拿来设计神经网络结构，防止多loss之间的balance，且是一个硬注入的知识

Thanks