

# **Authenticated and stable agentic communications**

Autonomous agents...controllers, bots, and AI agents...  
will dominate internet traffic and operation of other infrastructure

Prof. Alex Pentland, MIT and Stanford

# Example: Autonomous algorithms in trading

- AI Agents making complex, high-speed trades
- AI on distributed ledgers (blockchain): Swift, BRICS, Stablecoins
- Tokenization and other novel financial instruments will extend this algorithmic trading to all asset classes

# Existing agentic examples are worrisome:

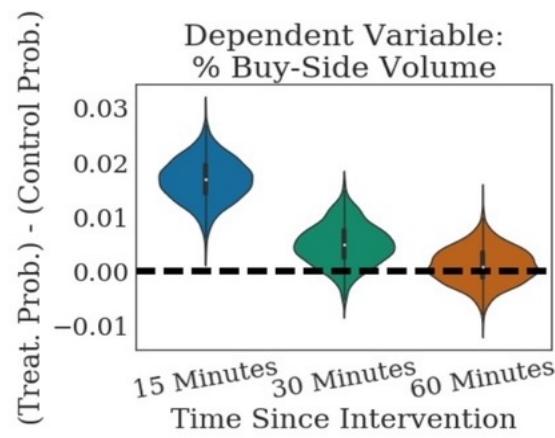
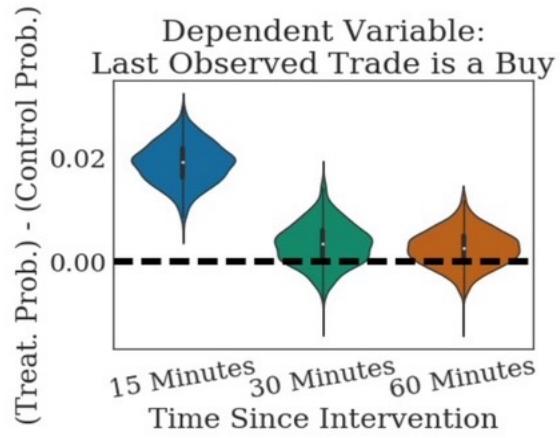
- Phantom traffic jams



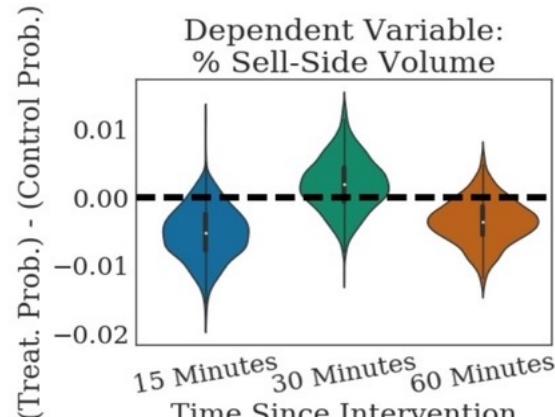
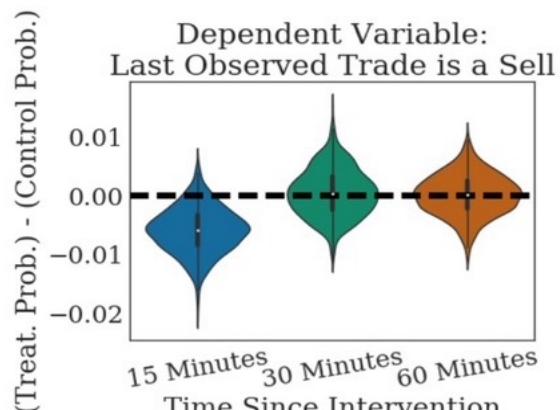
- Financial flash crashes



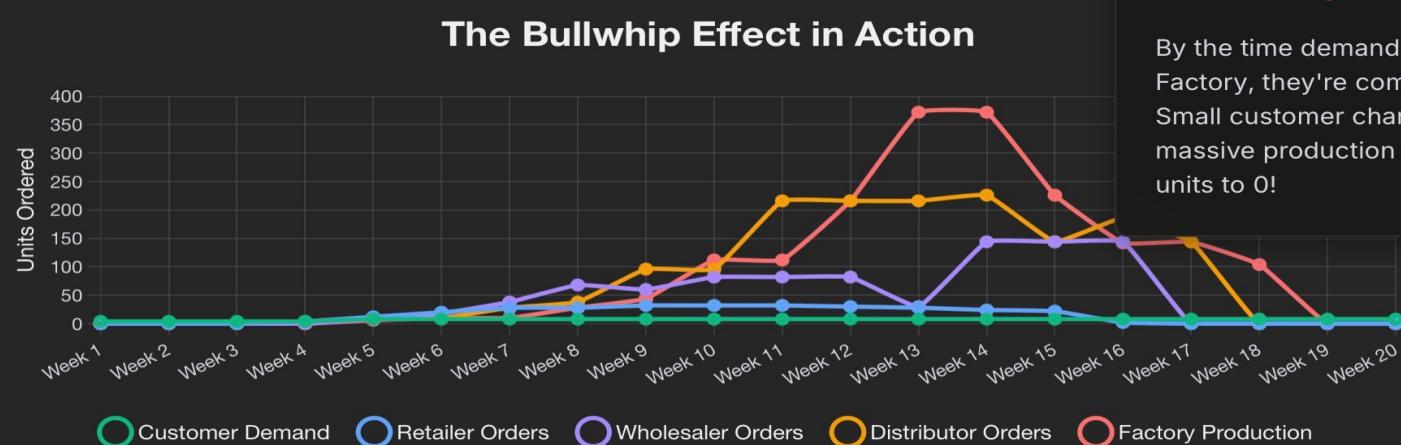
# High-speed, algorithmic crypto traders drive instability and prices



(a) Buy-side Interventions

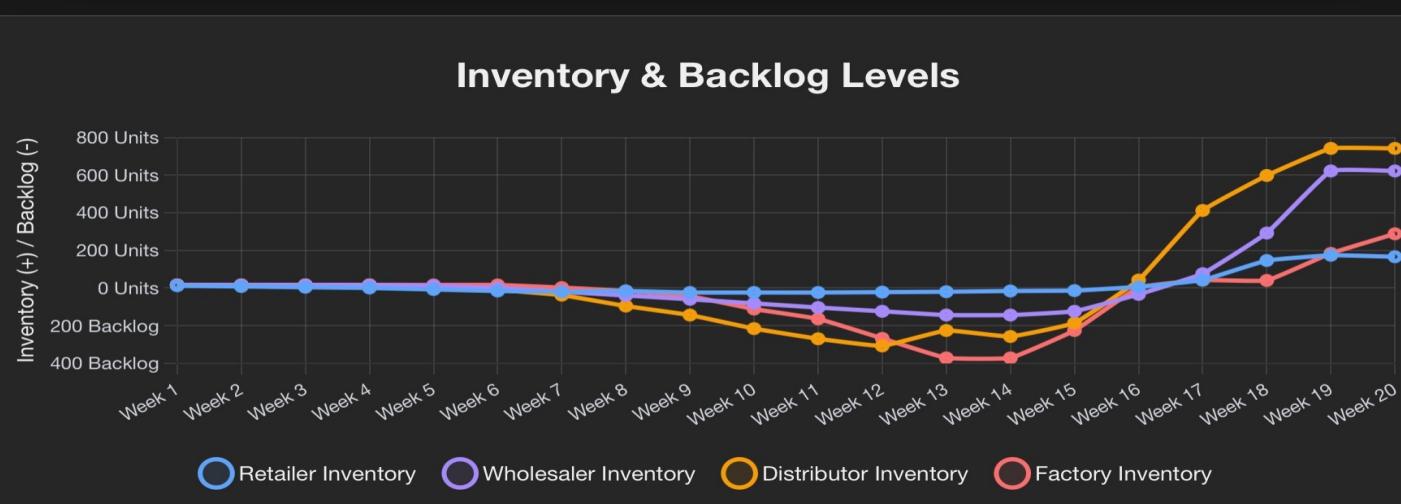


# Even simple networks of autonomous agents exhibit serious problems



### The Factory Extremes

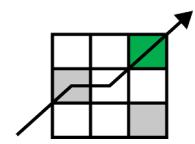
By the time demand signals reach the Factory, they're completely distorted. Small customer changes create massive production swings - from 372 units to 0!



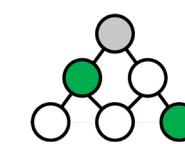
# And now, autonomous agents in every business process...e.g., Salesforce AgentForce, and SWIFT



User AI agents will  
**handle complex,  
laborious tasks**



Data flows through  
interfaces **controlled  
by user agents**,



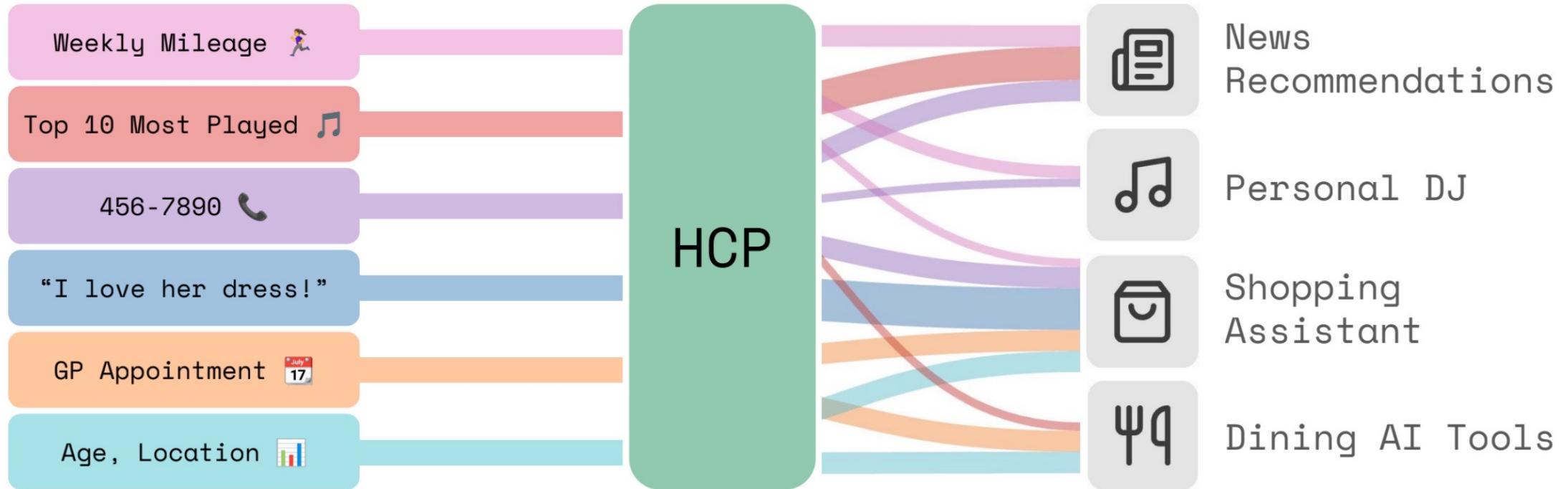
Companies are  
developing **agent  
APIs to directly  
connect to user  
agents**,

*h/t Jeremiah Ouyang*

## How can we insure a safe infrastructure?

# Possible Technical “fixes”

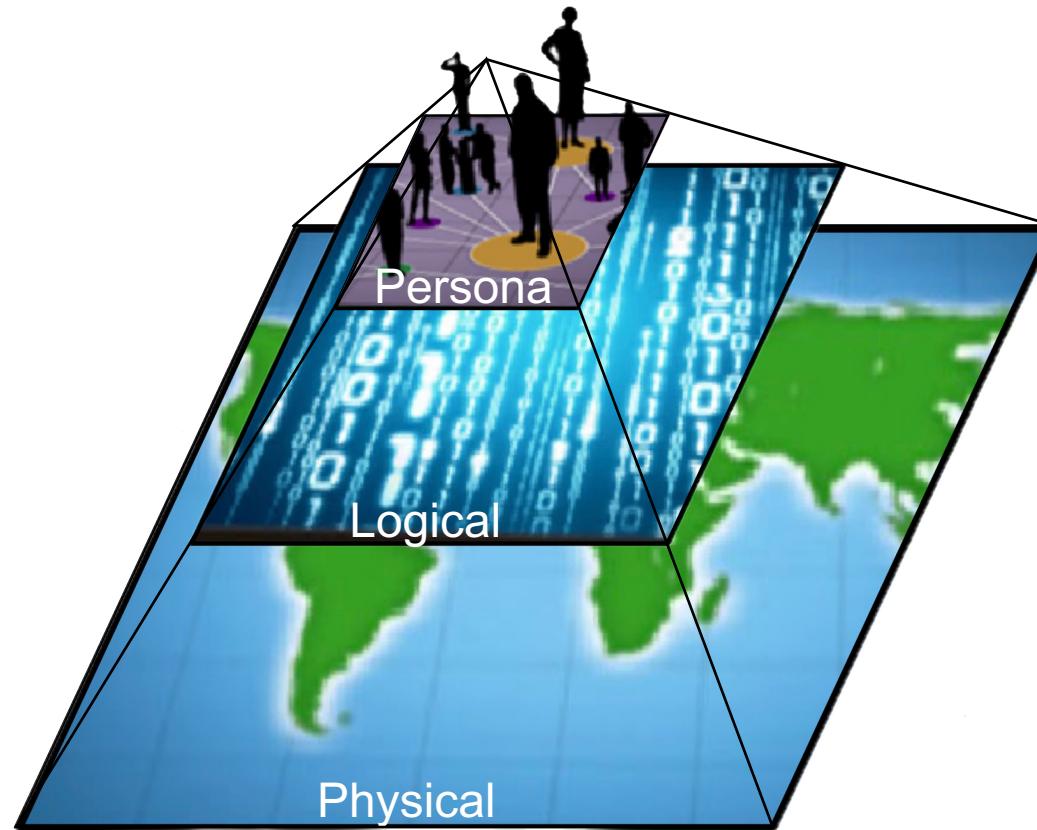
## Messaging with specification of authenticated actions



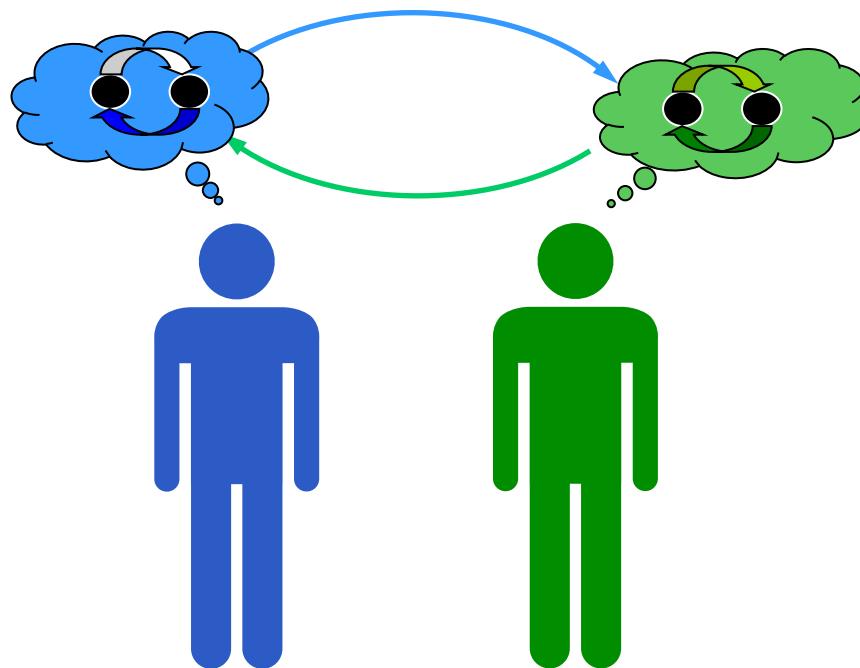
For example, specification of both  
recipient *and* response recipients

# Potential “fix”

## Identify unexpected coordination



# Unexpected Coordination Generates Unlikely Statistics

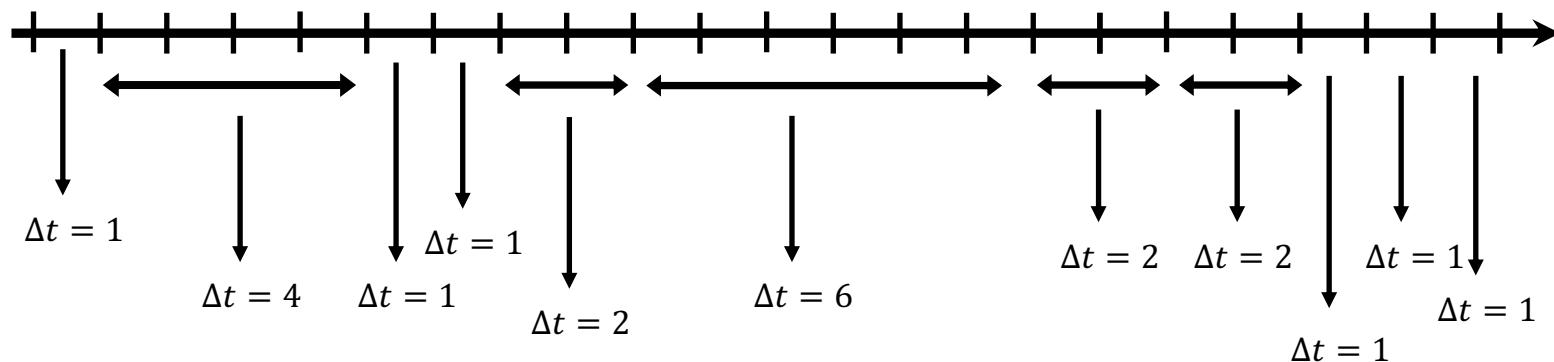


**Modeling Dynamical Influence in  
Human Interaction IEEE Signal Proc.**

$$\text{Prob}(h_t^{(c')} | h_{t-1}^{(1)}, \dots, h_{t-1}^{(C)}) = \sum_{c \in \{1, \dots, C\}} \underbrace{\mathbf{R}_{c',c}}_{\text{tie strength}} \times \underbrace{\text{Prob}(h_t^{(c')} | h_{t-1}^{(c)})}_{\text{cond. probability}}$$

# Real-time detection

The time difference between an individual's consecutive activities.

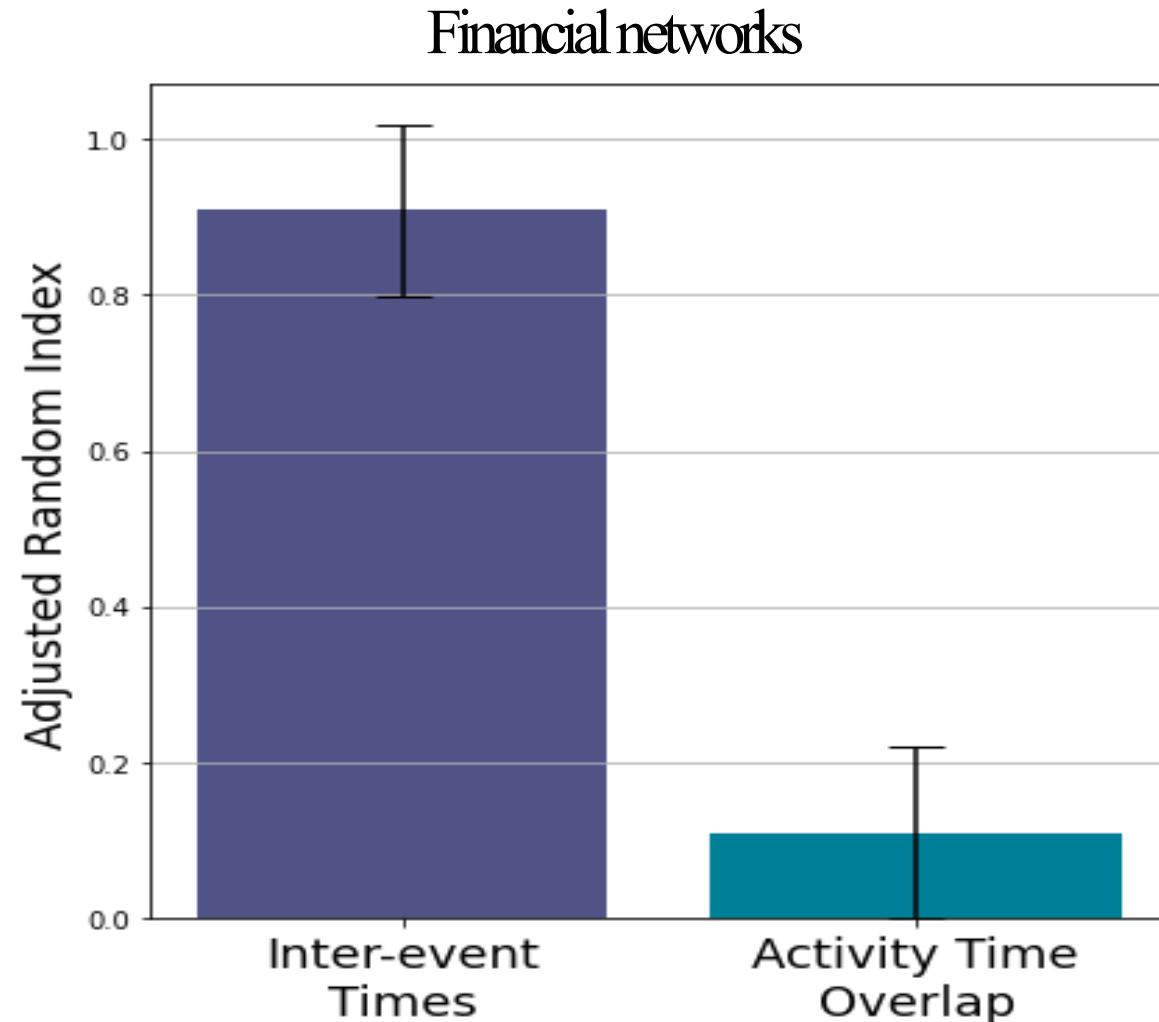
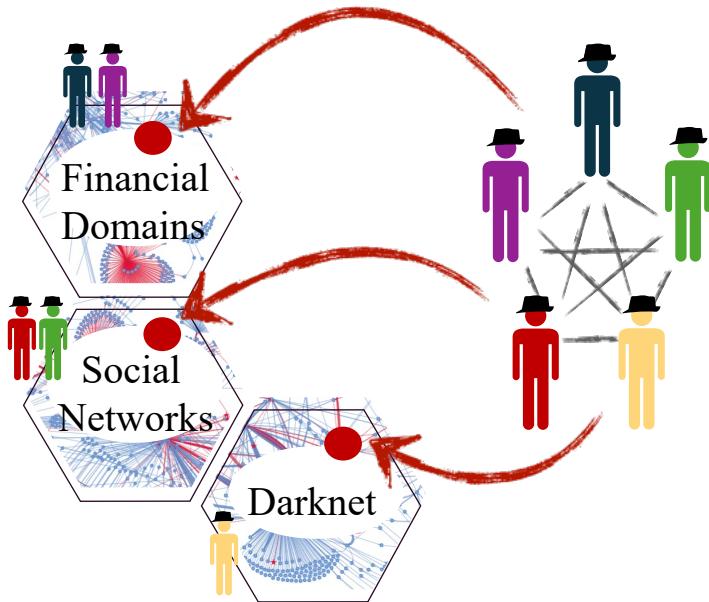


$$v_{\Delta t} = (1, 1, 1, 1, 1, 1, 2, 2, 2, 4, 6)$$

**Only need origin-destination matrix with time stamps**

# Coordinated Activity Detection even within fully encrypted domains!

Community detection  
across encrypted domains



## 1. Data:

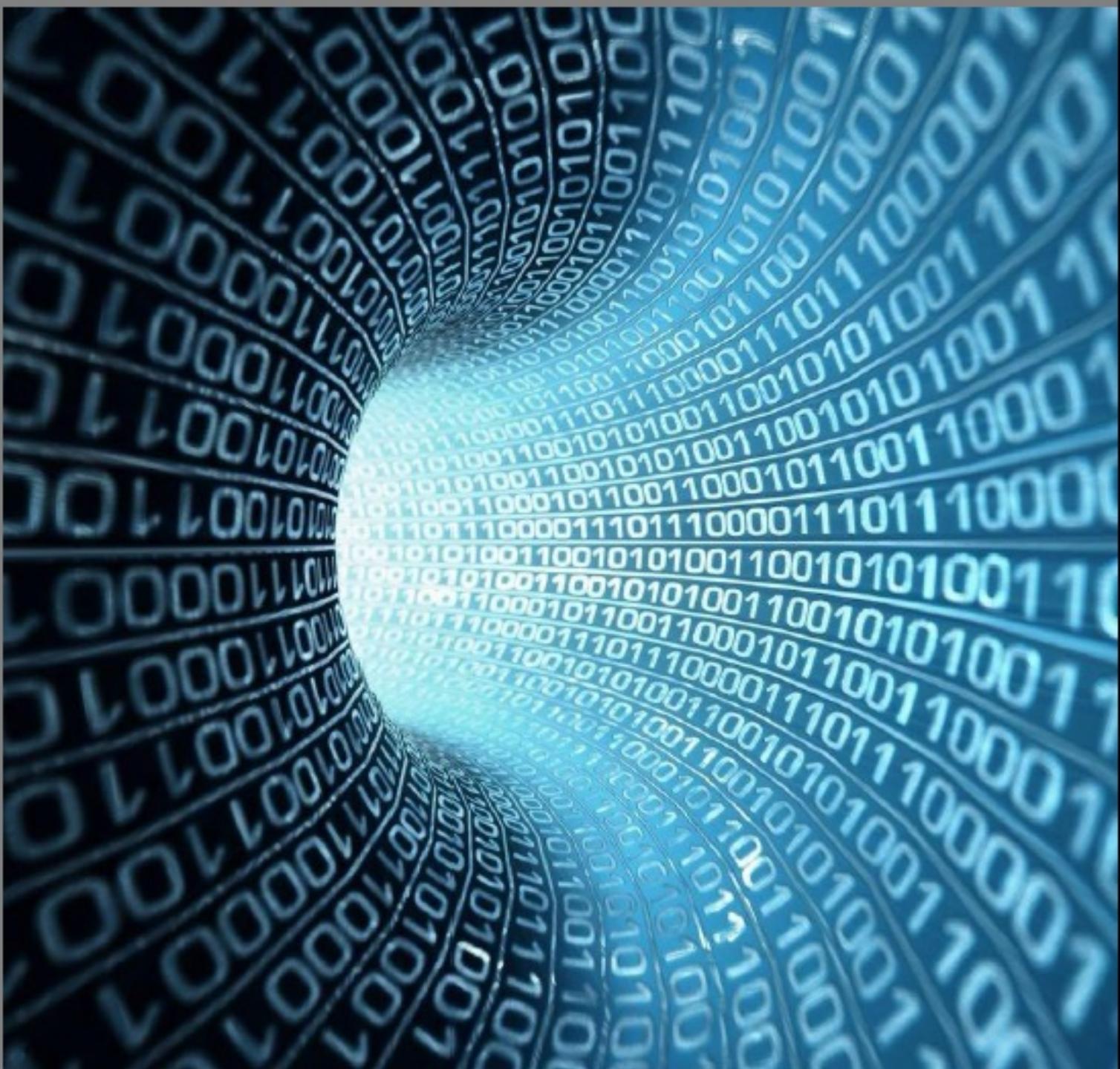
- Millions of Raw Twitter Tweets
- A list of 50 known ISIS Twitter accounts given as “sample”
- Additional 74 known ISIS accounts, kept hidden as test

## 2. Output:

- A list of “top 200 accounts most likely to be ISIS members”
- Runtime: 2 hours

## 3. Accuracy:

- 35 new ISIS accounts in top-50
- 51 new ISIS accounts in top-100
- 72 new ISIS accounts in top-200



# Summary:

- High-speed, autonomous agents pose a real threat to internet and infrastructure stability.
- Simple internet protocol changes could substantially limit the damage.