

# Decentralized Ensemble Learning and Online Social Networks

Presented by: Amira Soliman  
Supervisor: Šarūnas Girdzijauskas

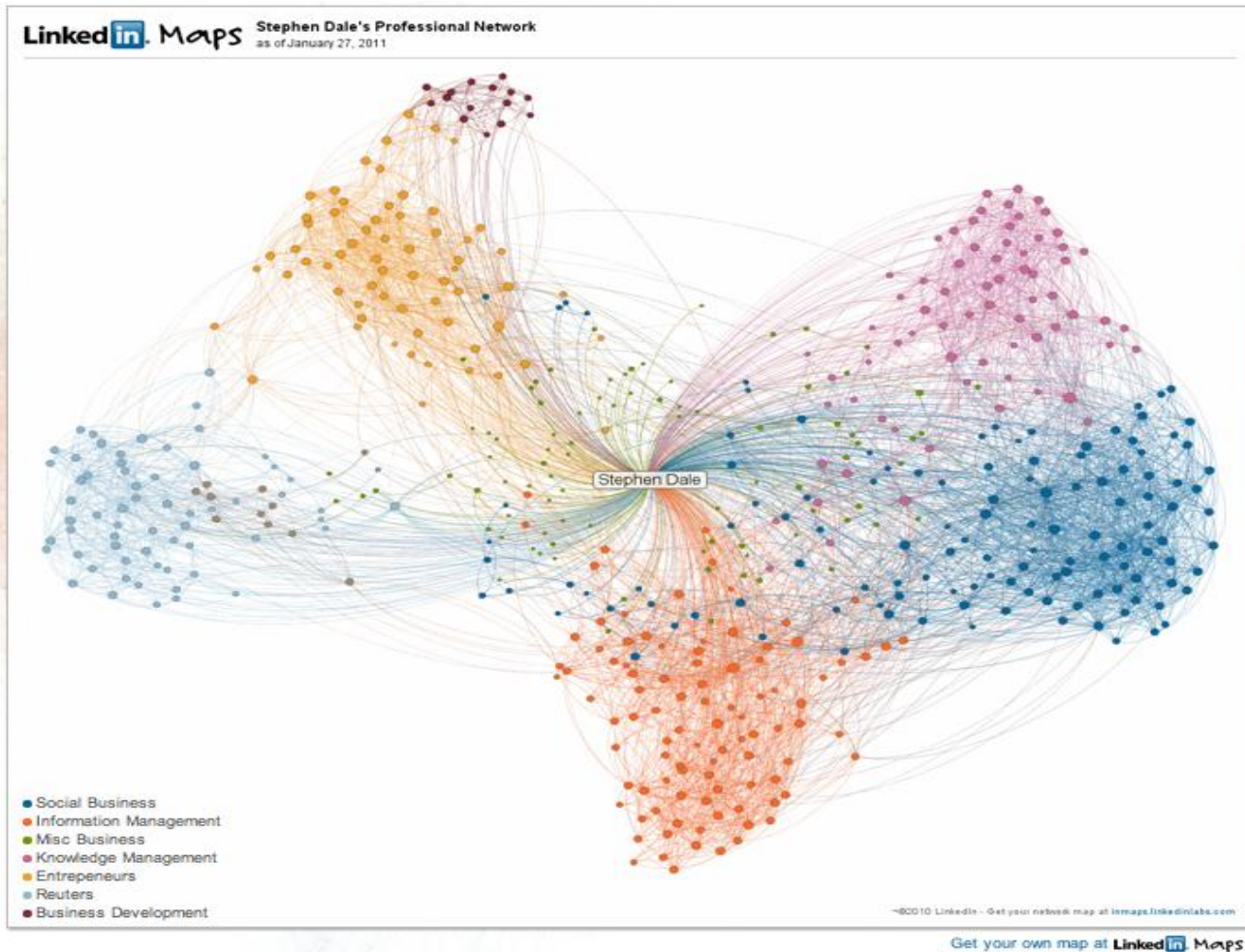
iSocial Meeting  
February 4-5<sup>th</sup>, 2014



# Outline

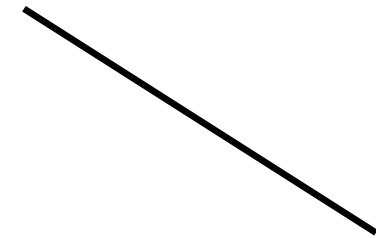
1. Distributed Machine Learning - Ensemble Learning
2. Gossip Learning on DOSNs
3. Sampling Service in DOSNs
4. Ensemble learners CTR model
5. Collaboration (Insubria & FORTH)

# Big Data & OSN



# Learn from Data - Model

Engine Simulation Program Features <i>39 Power Features™</i> (Numbered Items) Described Below	DeskTop And DynoSim Engine Simulations				Sport-Compact Engine Simulations			Dynamotion Professional Series Engine Simulations	
	DeskTop Dyno Basic	DeskTop Dyno	DynoSim Advanced	DynoSim ProTools™	DeskTop SC Dyno	SCDynoSim Advanced	SCDynoSim ProTools™	Dynamotion Advanced	Dynamotion ProTools™
Dyno-Testing RPM Range	2500 to 8000 rpm	2000 to 8500 rpm	1500 to 11500 rpm	1000 to 14500 rpm	2000 to 8500 rpm	1500 to 11500 rpm	1000 to 14500 rpm	1500 to 11500 rpm	1000 to 14500 rpm
Bore Range Limits	3.00 to 7.00-in	3.00 to 7.00-in	2.50 to 7.00-in	2.00 to 7.00-in	3.00 to 7.00-in	2.50 to 7.00-in	2.00 to 7.00-in	2.50 to 7.00-in	2.00 to 7.00-in
Stroke Range Limits	3.00 to 7.00-in	2.50 to 7.00-in	2.00 to 7.00-in	1.50 to 7.00-in	2.50 to 7.00-in	2.00 to 7.00-in	1.50 to 7.00-in	2.00 to 7.00-in	1.50 to 7.00-in
Includes Color Users Manual	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Best For Domestic Engines	Yes	Yes	Yes	Yes	---	---	---	Yes	Yes
Best For Sport Compact	---	---	---	---	Yes	Yes	Yes	Yes	Yes
Alternate Fuels/Nitrous (1)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
AirFlow Converter™ (2)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CamMath QuickCalculator™ (3)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Variable Valve Timing (4)	---	---	---	---	Yes	Yes	Yes	Yes	Yes
Advanced Compression-Ratio Calculator (5)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Connecting-Rod-Length And Rod-Ratio Modeling (6)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Forced-Induction Modeling (7)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
High-Speed Simulation (8)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Custom Cylinder-Head Flow (9)	No	---	---	---	Yes	Yes	Yes	Yes	Yes
Multi-Page Test Reports (10)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DirectClick™ Menus (11)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Real-Time "What-If" Testing (12)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
U.S./Metric Units (13)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes



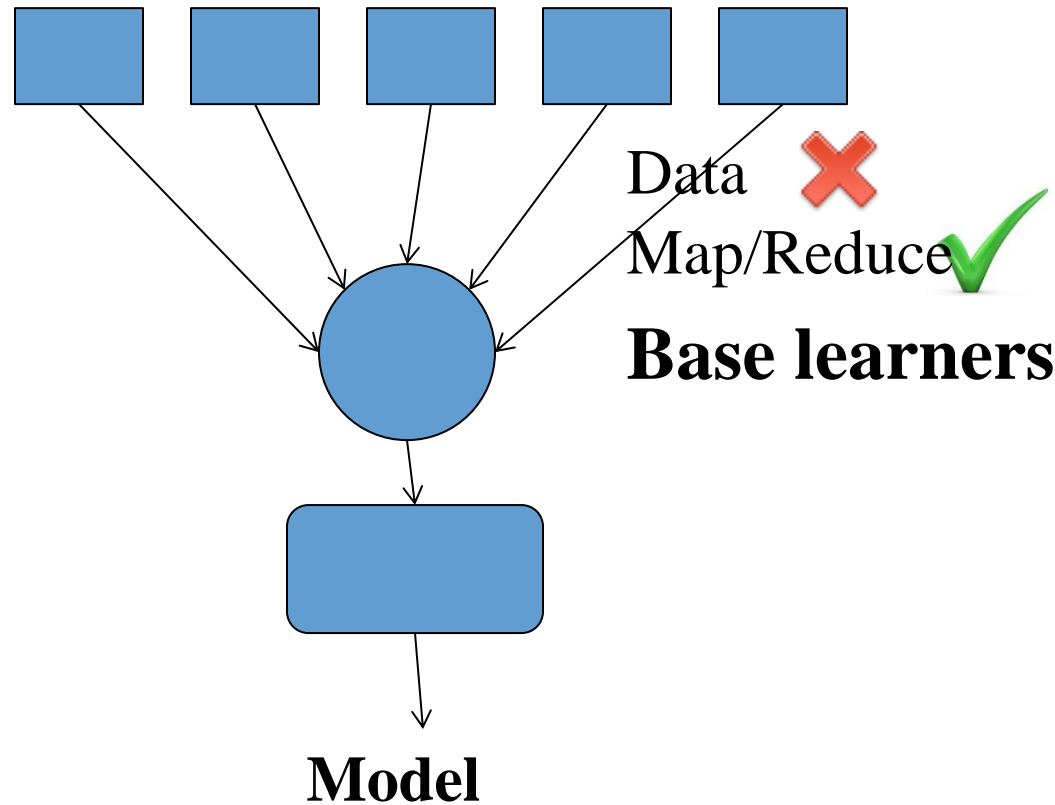
# Distributed Machine Learning

## Large-Scale Learning

- Data
- Time

## Gain?

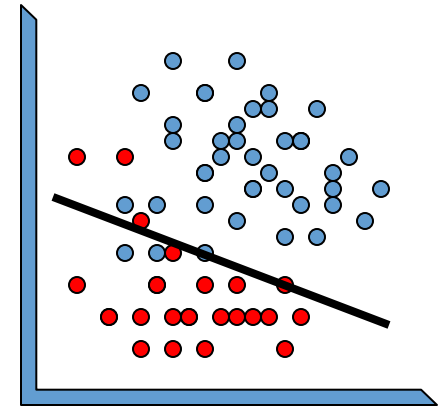
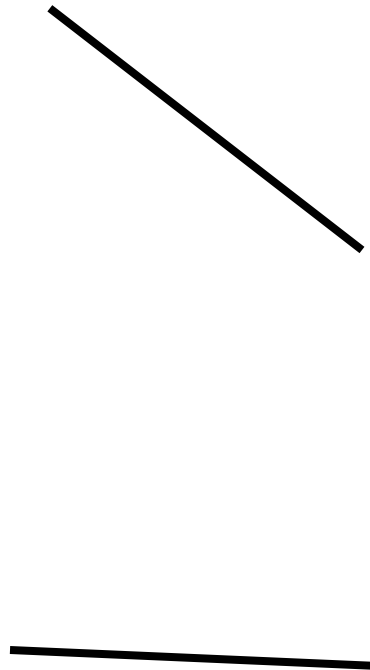
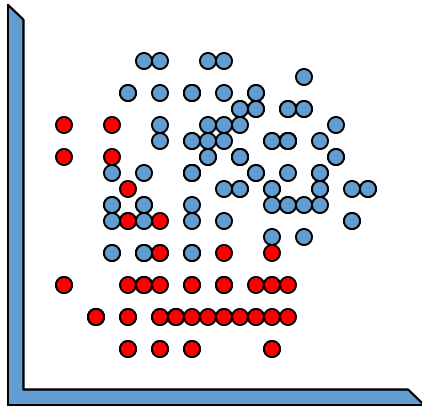
1. Computations
2. Network
3. Accuracy
4. Privacy



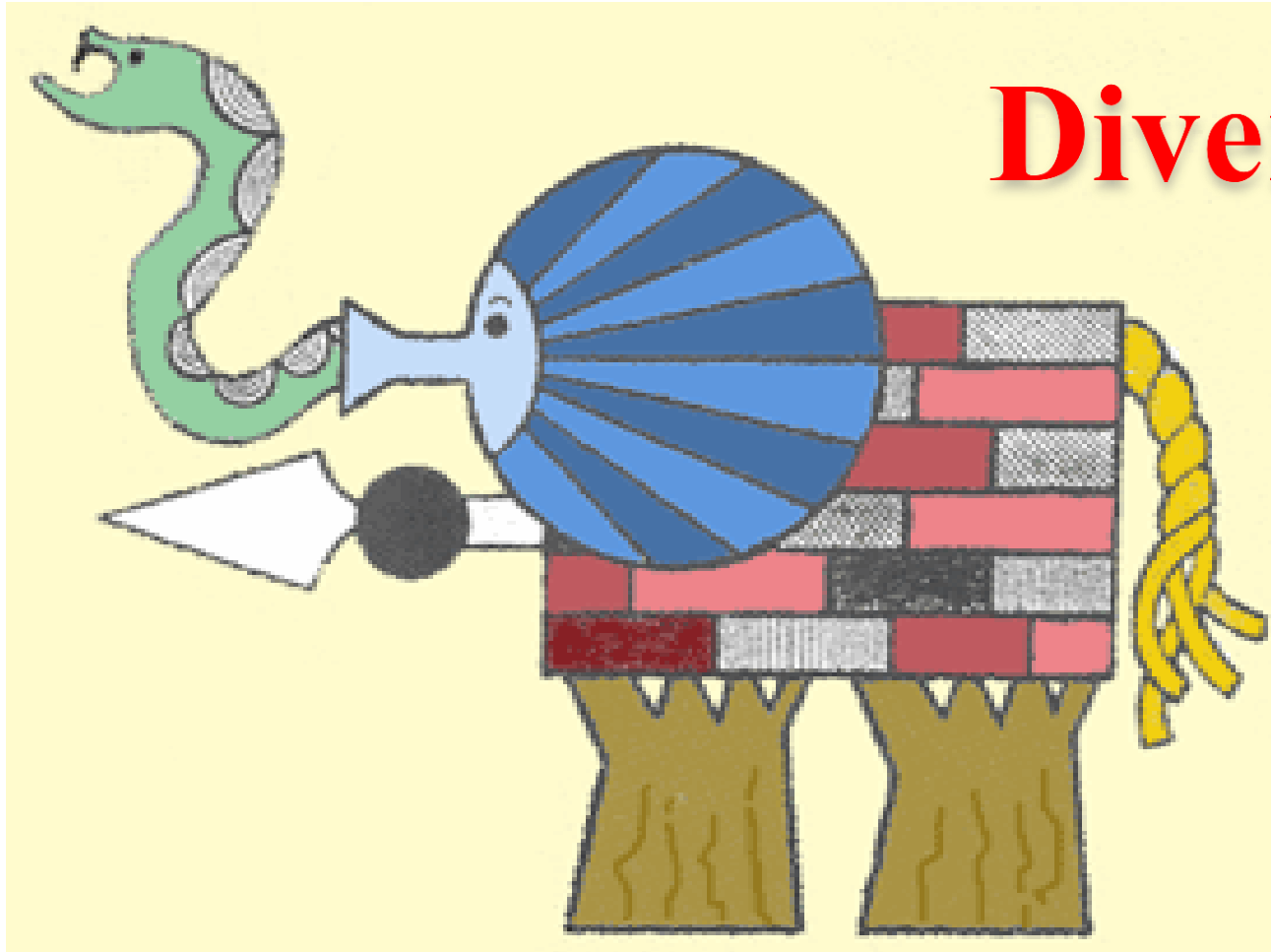
# Ensemble Learning

Two key Components:

- Algorithm to generate *Local Hypotheses*
- Method for combining output of these *Local Hypotheses*
  - *Predictions*
  - *Learners*



# Decentralized Ensemble Learning



**Diversity**

# Ensemble Learning on DOSNs

Why DOSN needs EL?

**Self-adaptive**

Ranking

Recommendation

Spam filtering

Community detection





# Ensemble Learning on DOSN

## Research Questions:

1. Diversity – random overlay
2. Navigation – constructing relay paths through social ties
3. Merge classifiers – reflect Ego/Socio – centric properties in OSN

## Methodologies:

1. Gossip-based Peer Sampling Service
2. On-the-fly Path Construction
3. Content, Time, and Reachability: CTR Model for merging classifiers

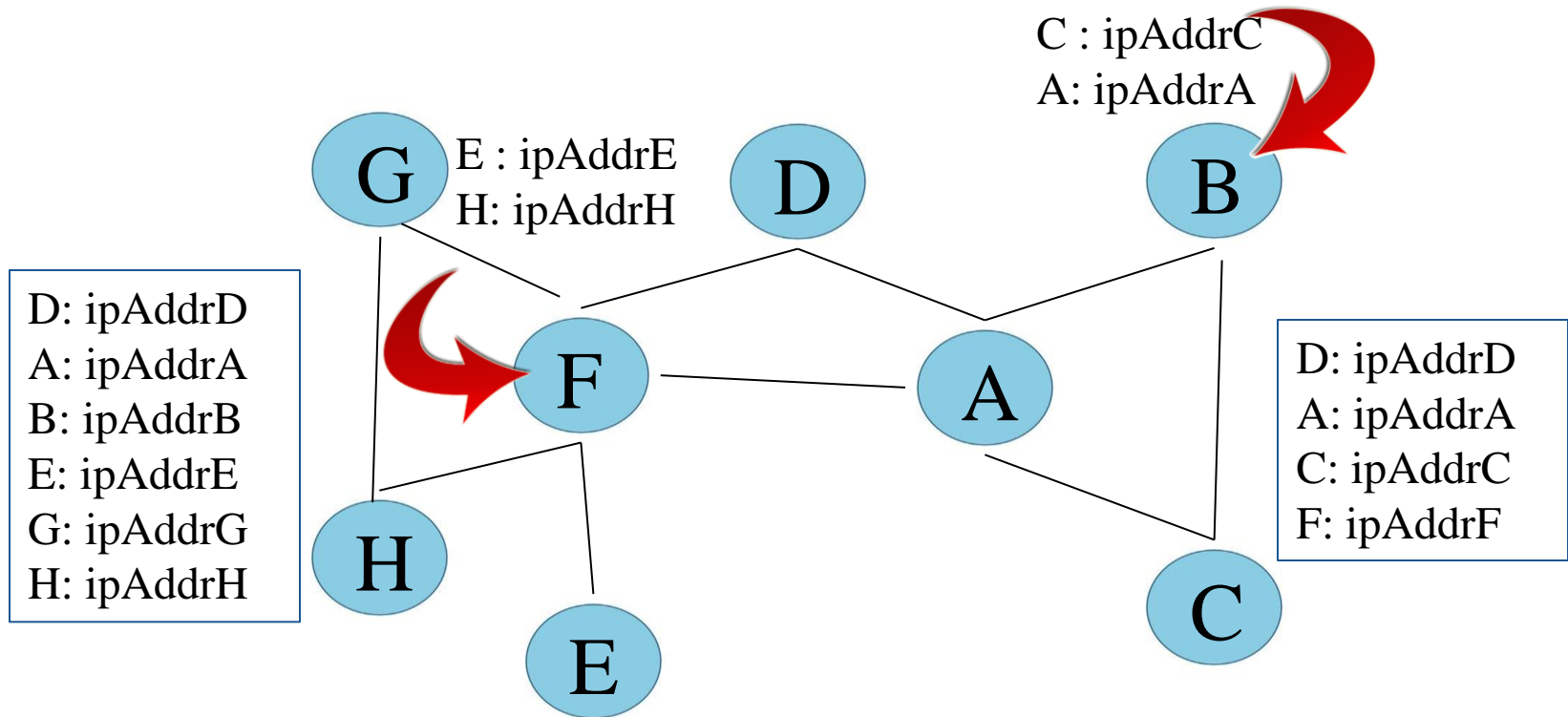
# Overlay Construction

Cyclon: Inexpensive membership management for unstructured P2P overlays

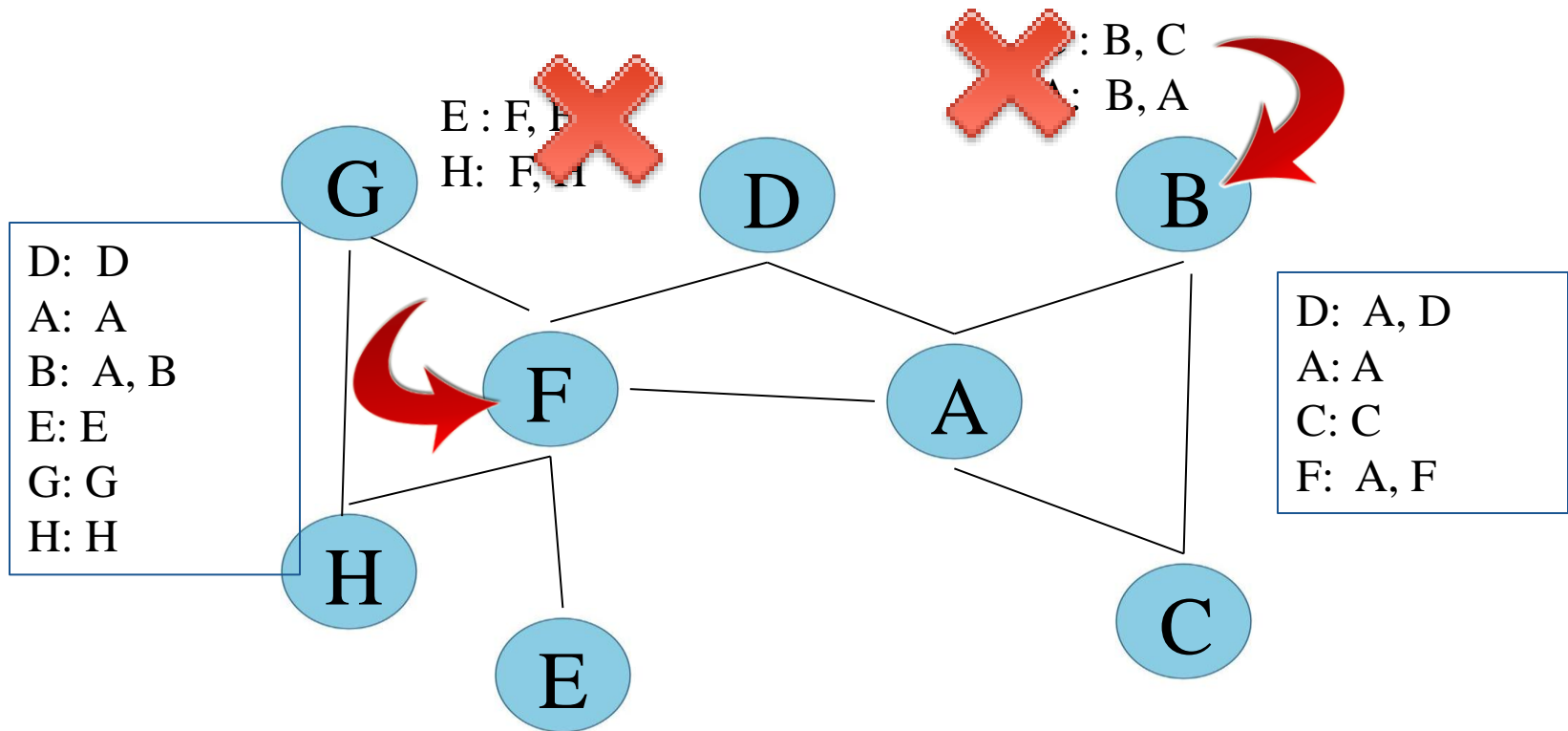
- that gives access to random peers
- has low diameter
- has low clustering coefficient
- Resilient to massive node failures

Basic idea: **Shuffle** operation, that's performed periodically using gossip

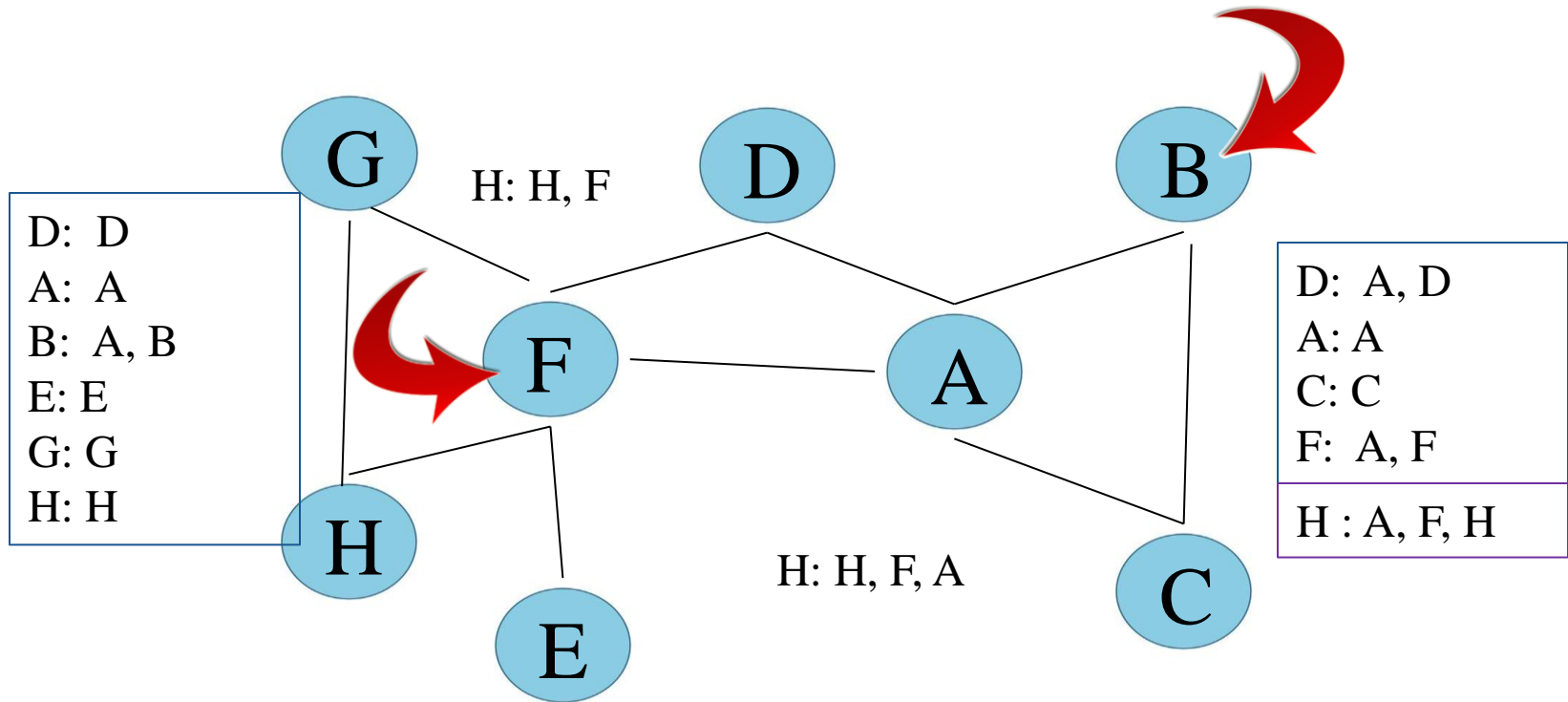
# How Cyclon works



# Overlay Construction in DOSN

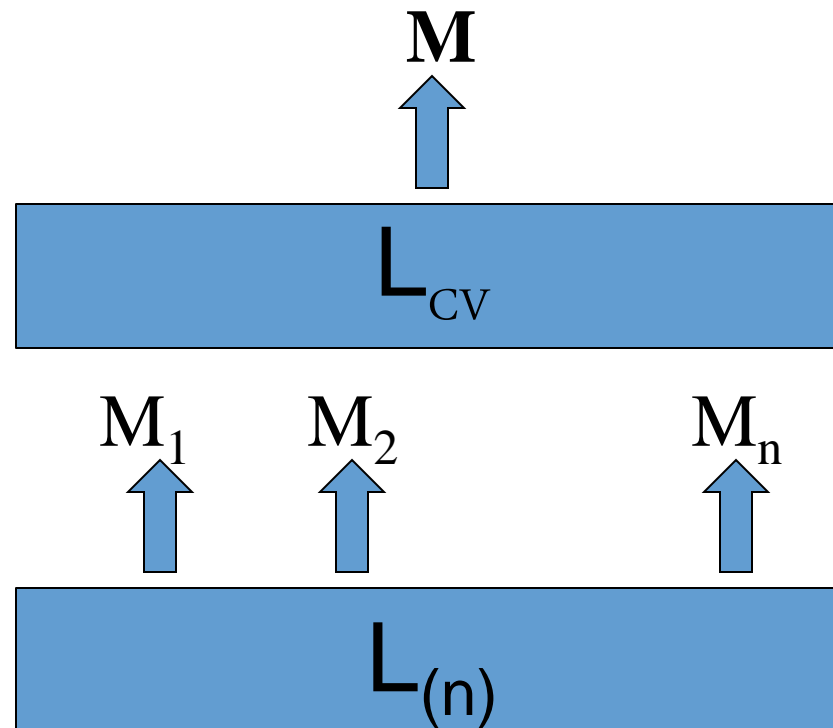


# On-the-fly Path Reconstruction



# Adaptive CTR Model

Stacking: training a learning algorithm to combine the predictions of several other learning algorithms.



- Content
- Time
- Reachability

# Applications

For our iSocial project:

1. Identity Management with INSUBRIA
2. Spam Filtering with FORTH

# Decentralized Privacy Management

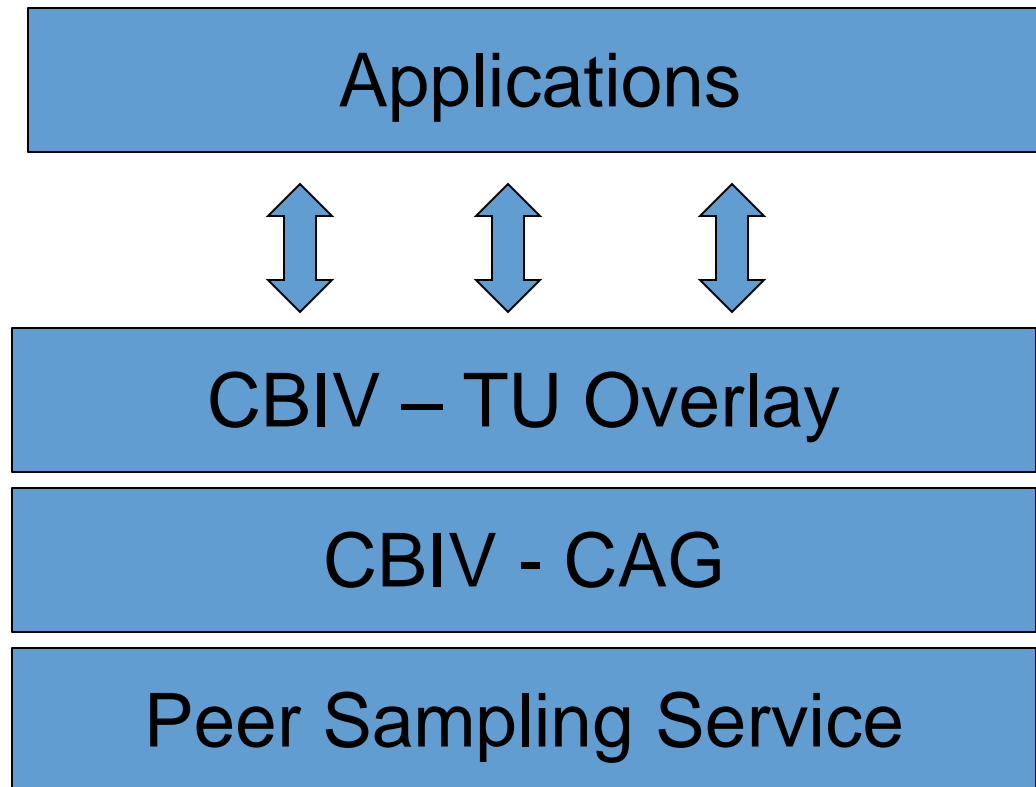
## Community-based Identity Validation:

1. Correlated Attribute Groups:
  - Find coherence relations
2. Raters Group: set of trusted users
  - Get feedback regarding new user profiles

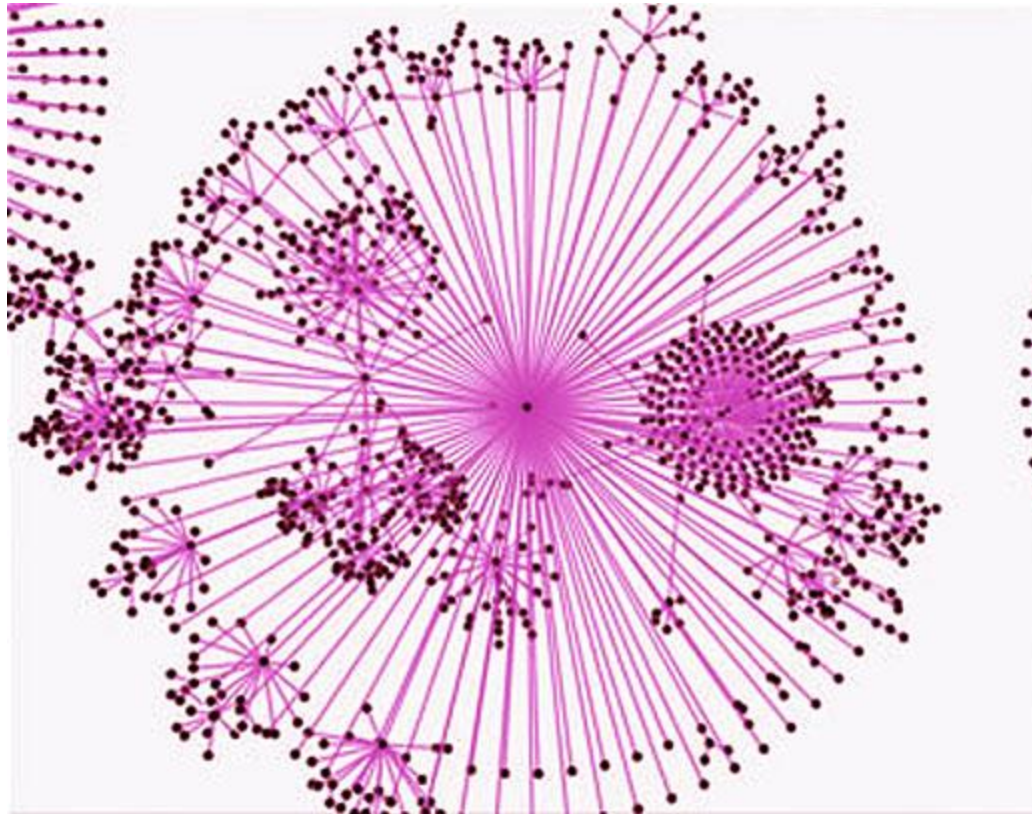




# Decentralized Privacy Management



# Decentralized Spam Filtering



# To Sum Up:

1. Ensemble Learning: use multiple models to obtain better predictive performance
2. Integrating Gossip protocols to develop decentralized ensemble learning strategy for DOSN (peer sampling service)
3. Stacked Generalization: adaptive CTR
4. Future work: decentralized privacy management and spam detection

That's all, thanks 😊

