



Ensemble Learning for Online Social Networks

Amira Soliman

PhD student, KTH/EES/LCN

Supervisor: Ass. Prof. Sarunas Girdzijauskas



Outline

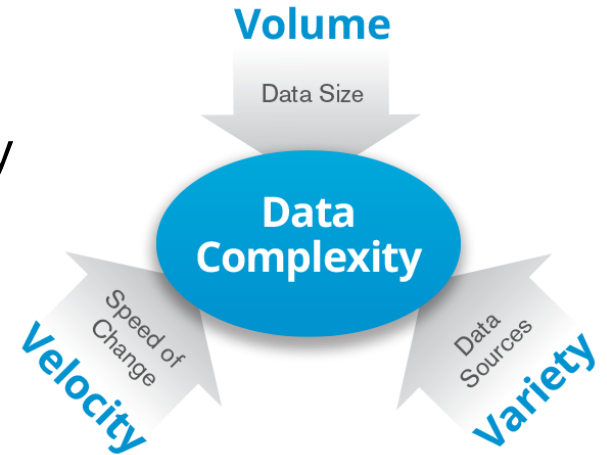
1. Ensemble Learning
2. Ensemble Learning for OSN
3. Peer Sampling Service
4. Merging Classifiers
5. Applications

Big Data Mining

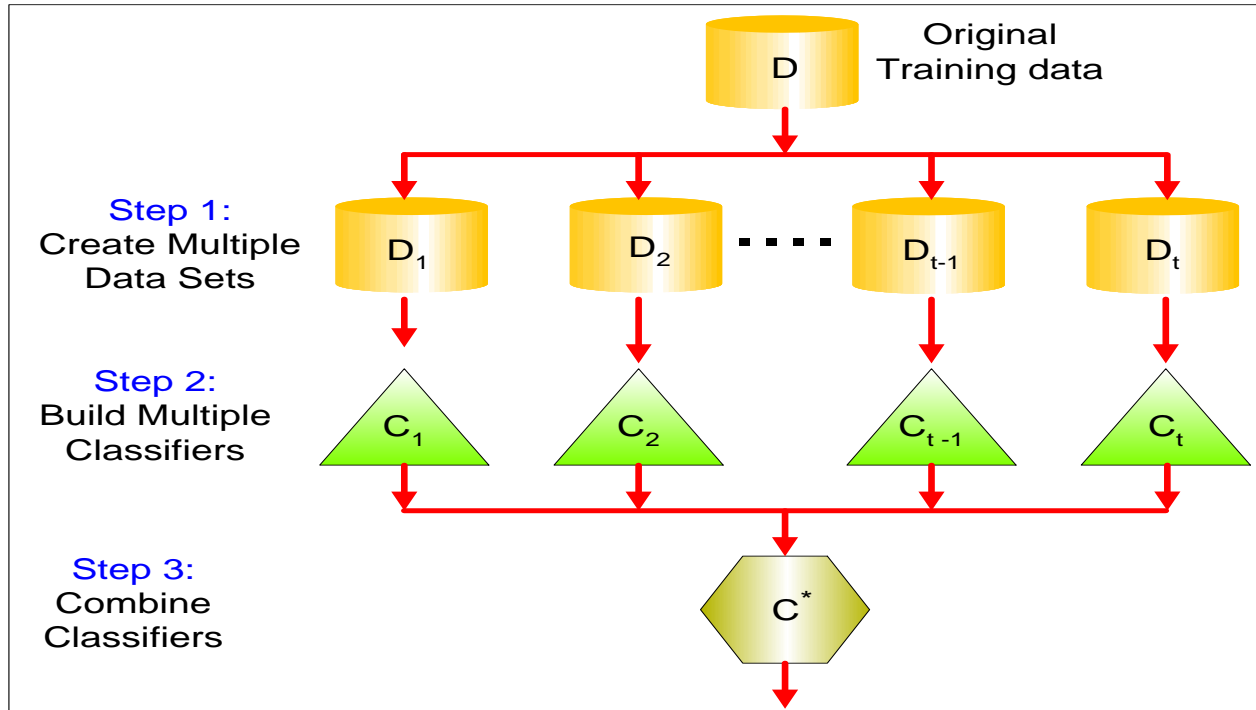
Turning this big data into “actionable insights” in a timely fashion

Challenges:

- Big Data 3V
- Information sharing and data privacy
- Mining complex and dynamic data
- Mining from sparse, uncertain, and incomplete data



Ensemble Learning





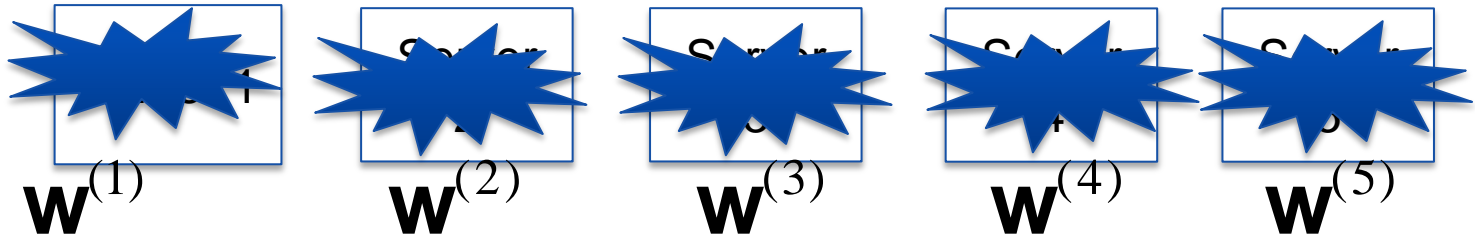
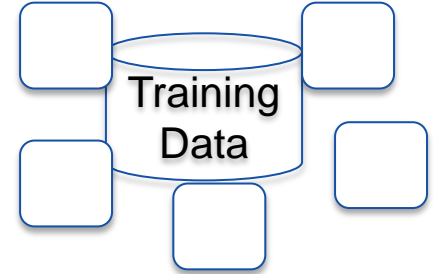
Motivations of Ensemble Learning

- Ensemble model improves accuracy and robustness over single model methods
- Applications:
 - distributed computing
 - privacy-preserving applications
 - large-scale data with reusable models
 - multiple sources of data

Simple Distributed Training

PerceptronParamMix($\mathcal{T} = \{(\mathbf{x}_t, \mathbf{y}_t)\}_{t=1}^{|\mathcal{T}|}$)

1. Shard \mathcal{T} into S pieces $\mathcal{T} = \{\mathcal{T}_1, \dots, \mathcal{T}_S\}$
2. $\mathbf{w}^{(i)} = \text{Perceptron}(\mathcal{T}_i)$
3. $\mathbf{w} = \sum_i \mu_i \mathbf{w}^{(i)}$
4. return \mathbf{w}



$$\mathbf{w} = \sum_i \mu_i \mathbf{w}^{(i)}$$

μ_i is the fraction of data for piece i

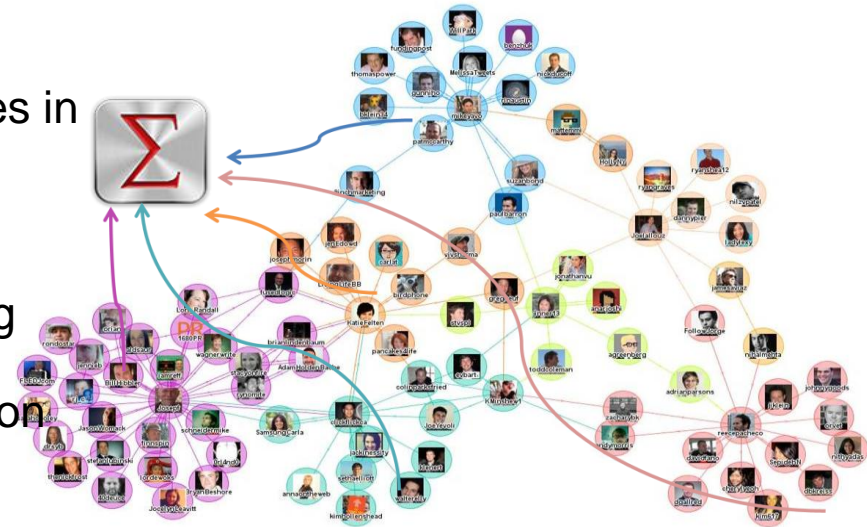
Ensemble Learning for OSN

Research Questions:

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

Methodologies:

- Gossip-based Peer Sampling Service
- On-the-fly Path Reconstruction
- Merging classifiers based on content-richness.





Our approach

1. Construct a random overlay on top of the network.
2. Train classifier with local training dataset.
3. Receive classifiers from neighbors in overlay.
4. Assign the weights of different classifiers.
5. Update local classifier with the using the weighted average:

$$w = \sum_{k=1}^N \alpha_k \cdot w_k$$

6. Propagate updated classifier to neighbors in overlay.



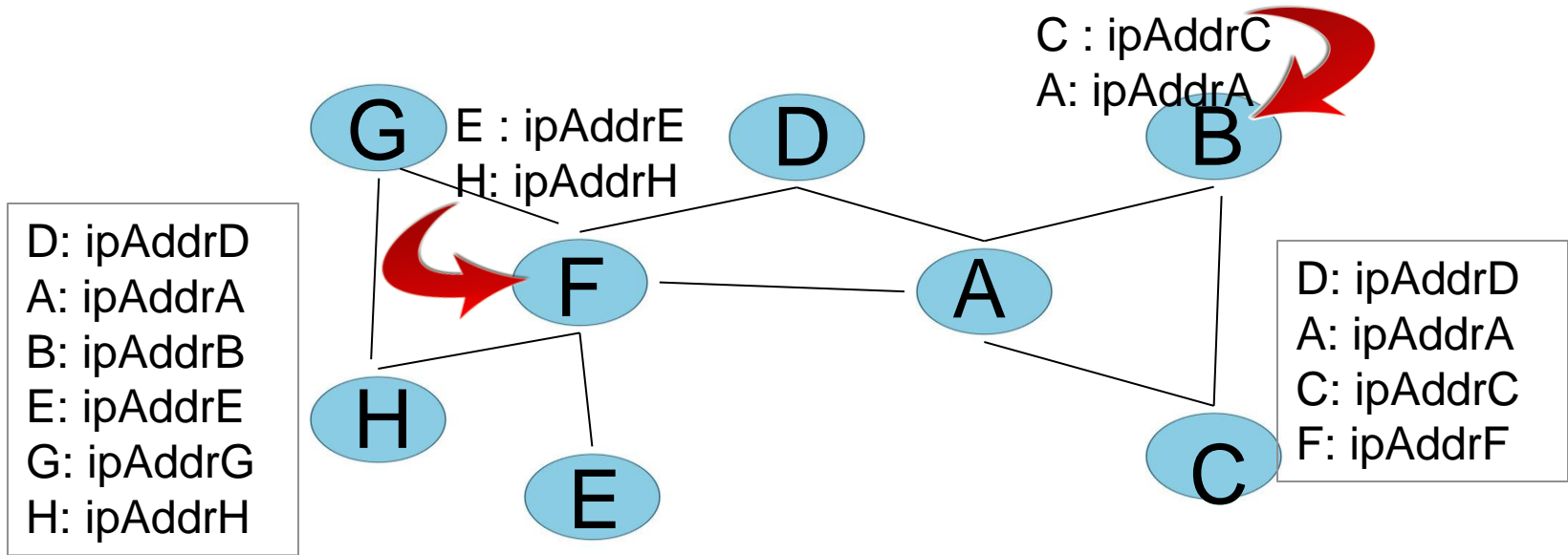
Overlay Construction

Cyclon: Inexpensive membership management for unstructured P2P overlays

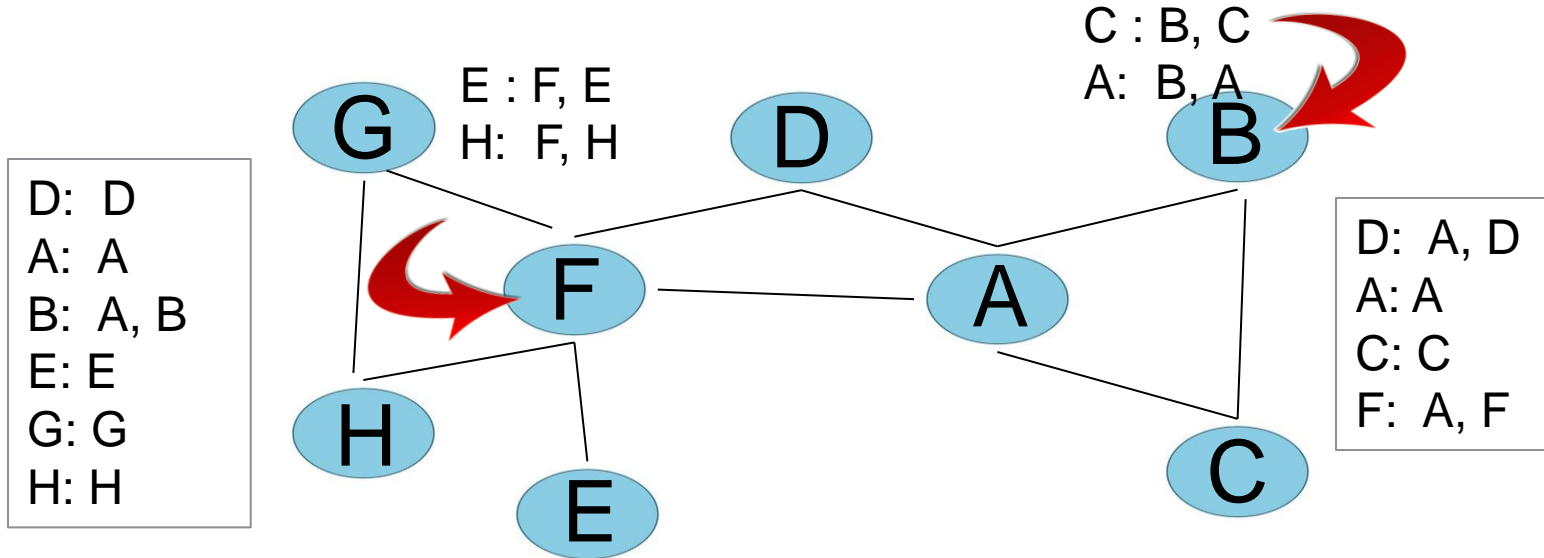
- that gives access to random peers
- has low diameter
- has low clustering coefficient
- Resilient to massive node failures

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

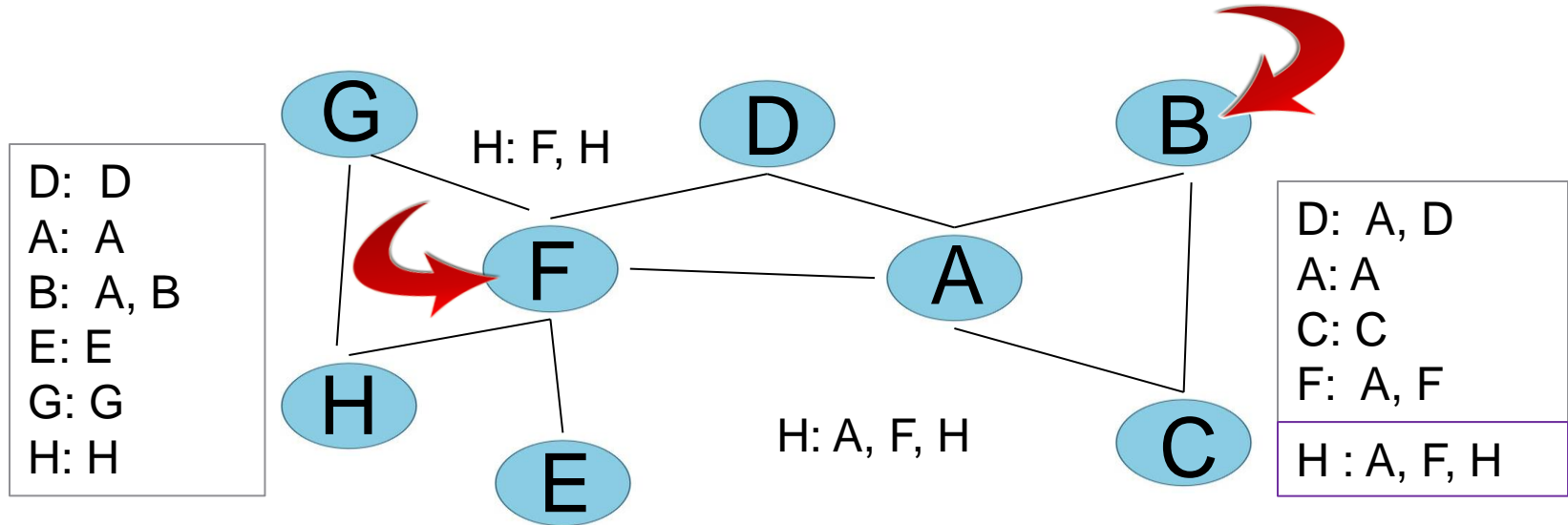
How Cyclon works



Overlay Construction on top of OSN



On-the-fly Path Reconstruction





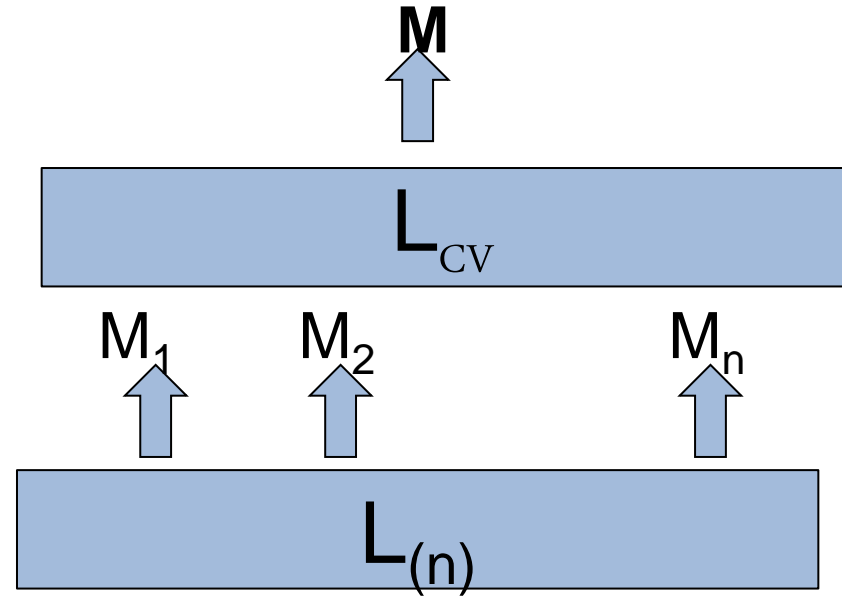
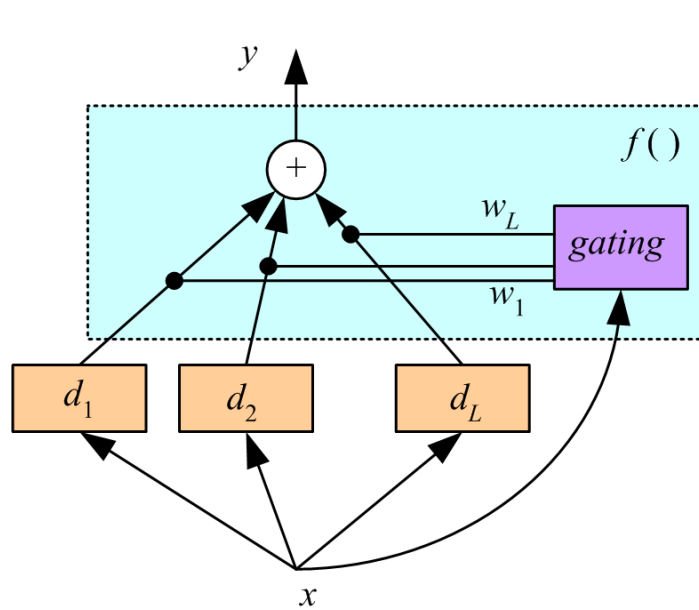
Our approach

1. Construct a random overlay on top of the network.
2. Train classifier with local training dataset.
3. Receive classifiers from neighbors in overlay.
4. Assign the weights of different classifiers.
5. Update local classifier with the using the weighted average:

$$w = \sum_{k=1}^N \alpha_k \cdot w_k$$

6. Propagate updated classifier to neighbors in overlay.

Merging Models





Evaluation

1. We have generated synthetic graphs and data.
2. Data from OSN (Facebook, and Twitter)

Applications:

For our iSocial project:

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



Thanks

Questions and Comments are All welcome 😊